```
                    GenCore version 4.5
            Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          May 31, 2002, 15:47:13 ; Search time 15.22 Seconds
                                           (without alignments)
                                           2599.959 Million cell updates/sec

Title:           US-09-049-304A-122
Perfect score:   5286
Sequence:        1 CARLLLGGGKNGPRVNRIIV..........VPALEILESSGIKLVEKVET 1022

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters:       105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_40:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES

                     %
Result             Query
  No.     Score    Match Length DB  ID                       Description
--------------------------------------------------------------------------------
    1     672.5     12.7    446  1  LYS9_YEAST               P38999 saccharomyc
    2       172      3.3    369  1  LYS1_YARLI               P38997 yarrowia li
    3       157      3.0    368  1  LYS1_SCHPO               Q09694 schizosacch
    4     153.5      2.9    373  1  LYS1_YEAST               P38998 saccharomyc
    5     144.5      2.7    382  1  LYS1_CANAL               P43065 candida alb
    6     141.5      2.7   1861  1  MAP2_RAT                 P15146 rattus norv
    7     139.5      2.6    604  1  DNAK_BACME               P05646 bacillus me
    8     125.5      2.4   1451  1  SPT6_YEAST               P23615 saccharomyc
    9       124      2.3   5255  1  BACA_BACLI               P23615 saccharomyc
   10       121      2.3   2672  1  GCN1_YEAST               O68006 b bacitraci
   11       120      2.3    493  1  CETP_HUMAN               P33892 saccharomyc
   12       119      2.3   1828  1  MAP2_MOUSE               P11597 homo sapien
   13       118      2.2    428  1  ENO_ZYMMO                P20357 mus musculu
   14     116.5      2.2   1450  1  RPO1_ASFB7               P33675 zymomonas m
   15     116.5      2.2   4644  1  DYHC_MOUSE               P42486 african swi
   16     115.5      2.2    493  1  AMYR_DROTE               Q9jhu4 mus musculu
   17     115.5      2.2    494  1  AMYR_DROBP               O76260 drosophila
   18     115.5      2.2    562  1  EZRA_BACSU               Q9njn8 drosophila
   19     115.5      2.2    680  1  LAG2_YEAST               O34894 bacillus su
   20     115.5      2.2   3587  1  TYCB_BACBR               Q92325 saccharomyc
   21       115      2.2    858  1  YNU1_YEAST               O30408 b tyrocidin
   22     114.5      2.2   3176  1  CA36_HUMAN               P40164 saccharomyc
   23       114      2.2    890  1  POL2_BAYMG               P12111 homo sapien
   24       114      2.2   1416  1  RPOC_TREPA               Q01365 barley yell
   25     113.5      2.1    986  1  GUNZ_CLOSR               O83270 treponema p
   26       113      2.1   1244  1  DPOL_ASFL6               P23659 clostridium
   27     112.5      2.1    313  1  DHCO_LACLA               P43139 african swi
   28     112.5      2.1    626  1  HCY6_ANDAU               P15244 lactococcus
   29     112.5      2.1   2863  1  LRBA_HUMAN               P80476 androctonus
   30       112      2.1   1862  1  ANK1_MOUSE               P50851 homo sapien
   31     111.5      2.1    493  1  AMYR_DROSE               Q02357 mus musculu
   32     111.5      2.1   1023  1  RT12_ACTPL               O76261 drosophila
   33     111.5      2.1   1679  1  YIO9_YEAST               P55129 actinobacil
                                                             P40457 saccharomyc
```

```
   34       111      2.1   1240  1  DP3A_CHLPN               Q9z7n8 chlamydia p
   35       110      2.1   1132  1  PHY1_PHYPA               P36505 physcomitre
   36     109.5      2.1    493  1  AMYR_DROER               O76265 drosophila
   37     109.5      2.1    493  1  AMYR_DROME               O18408 drosophila
   38     109.5      2.1    610  1  DNAK_BACSH               O69268 bacillus sp
   39     109.5      2.1   1595  1  SOS_DROME                P26675 drosophila
   40       109      2.1   1024  1  POPC_RALSO               Q9rbs2 ralstonia s
   41       109      2.1   1592  1  YNY2_YEAST               P53855 saccharomyc
   42     108.5      2.1    493  1  AMYR_DROYA               O76264 drosophila
   43     108.5      2.1    511  1  ATPB_YEAST               P00830 saccharomyc
   44     108.5      2.1    977  1  A2A1_MOUSE               P17426 mus musculu
   45     108.5      2.1    984  1  SECA_AQUAE               O67718 aquifex aeo

                                  ALIGNMENTS

RESULT   1
LYS9_YEAST
ID   LYS9_YEAST     STANDARD;      PRT;   446 AA.
AC   P38999;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Saccharopine dehydrogenase [NADP+, L-glutamate forming] (EC 1.5.1.10).
GN   LYS9 OR LYS13 OR YNR050C OR N3461.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1278B;
RA   Feller A.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Pohl T.M.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 314-324.
RC   STRAIN=ATCC 38531 / Y41;
RX   MEDLINE=97089742; PubMed=8935650;
RA   Norbeck J., Blomberg A.;
RT   "Protein expression during exponential growth in 0.7 M NaCl medium of
RT   Saccharomyces cerevisiae.";
RL   FEMS Microbiol. Lett. 137:1-8(1996).
CC   -!- CATALYTIC ACTIVITY: N6-(L-1,3-dicarboxypropyl)-L-lysine + NADP(+)
CC       + H(2)O = L-glutamate + 2-aminoadipate 6-semialdehyde + NADPH.
CC   -!- PATHWAY: SEVENTH STEP IN LYSINE BIOSYNTHESIS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X77363; CAA54552.1; -.
DR   EMBL; Z71665; CAA96331.1; -.
DR   PIR; S41937; S41937.
DR   YEPD; 8416; -.
DR   SGD; S0005333; LYS9.
KW   Lysine biosynthesis; Oxidoreductase; NADP.
SQ   SEQUENCE   446 AA;  48917 MW;  1EC4CE1CE3BCD916 CRC64;

  Query Match             12.7%;  Score 672.5;  DB 1;  Length 446;
  Best Local Similarity   35.1%;  Pred. No. 3.5e-36;
  Matches  170;  Conservative   82;  Mismatches  183;  Indels   49;  Gaps   13;

Qy        543 GPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGIEN 602
```

```
          |  :|:||:| | :|  : ||:              | |:| ||     :|:
Db      2 GKNVLLLGSGFVAQPVIDTLAA-------------NDDINVTVACRTLANAQALAKP-SG 47

Qy    603 TTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNL 662
          : |  ||| |  :|  :::  :|||||:| :||  :    |  |  :||:||:  ::  |
Db     48 SKAISLDVTDDSALDKVLADNDVVISLIPYTFHPNVVKSAIRTKTDVVTSSYISPALREL 107

Qy    663 SQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPL 722
                ||:|:: |:||||||||| ::| ||| |   ||:|:| ||||||||:|  ::|||
Db    108 EPEIVKAGITVMNEIGLDPGIDHLYAVKTIDEVHRAGGKLKSFLSYCGGLPAPEDSDNPL 167

Qy    723 AYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNS 782
           |||||:  | | : :| | |   |:   |   :|  :||   :   | :|    |||:|
Db    168 GYKFSWSSRGVLLALRNSAKYWKDGKIETVSSEDLMATAKPYFI--YPGYAFVCYPNRDS 225

Qy    783 LIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFF-DAANH----PLLQDTSRPTY 837
           :: ||| |  || |: | | ||:|| | :  |   |   | ||     |:  : :   |
Db    226 TLFKDLYHI-PEAETVIRGTLRYQGFPEFVKALVDMGMLKDDANEIFSKPIAWNEALKQY 284

Qy    838 KGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHE 897
           |                   |     :|||| :      |: |   : :    :|||
Db    285 LG-------------------AKSTSKEDLIASIDSKATWKDDEDRERILSGFAWLGLFS 325

Qy    898 ETQI-PKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEF 956
          : :| |:|  :  | :| |:|: | |  ||:|||:| |:  :|: ||   |   :||:::
Db    326 DAKITPRG--NALDTLCARLEELMQYEDNERDMVVLQHKFGIEWADG-TTETRTSTLVDY 382

Qy    957 GKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPAL-EILESSGIK 1015
          |||  | |  :|| ||| | ||    :|   ::  |::| |  ||| | : |: :  ||
Db    383 GKV-GGYS--SMAATVGYPVAIATKFVLDGTIKGPGLLAPYSPEINDPIMKELKDKYGIY 439

Qy   1016 LVEK 1019
          | ||
Db    440 LKEK 443

RESULT    2
LYS1_YARLI
ID   LYS1_YARLI      STANDARD;      PRT;   369 AA.
AC   P38997;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Saccharopine dehydrogenase [NAD+, L-lysine forming] (EC 1.5.1.7)
DE   (Lysine--2-oxoglutarate reductase) (SDH).
GN   LYS5.
OS   Yarrowia lipolytica (Candida lipolytica).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Dipodascaceae; Yarrowia.
OX   NCBI_TaxID=4952;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 20460 / W29;
RX   MEDLINE=90355996; PubMed=2388625;
RA   Xuan J.-W., Fournier P., Declerck N., Chasles M., Gaillardin C.;
RT   "Overlapping reading frames at the LYS5 locus in the yeast Yarrowia
RT   lipolytica.";
RL   Mol. Cell. Biol. 10:4795-4806(1990).
CC   -!- CATALYTIC ACTIVITY: N6-(L-1,3-dicarboxypropyl)-L-lysine + NAD(+) +
CC       H(2)O = L-lysine + 2-oxoglutarate + NADH.
CC   -!- PATHWAY: 8TH STEP IN LYSINE BIOSYNTHESIS.
CC   -!- SIMILARITY: TO OTHER FUNGAL SACCHAROPINE DEHYDROGENASE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as  long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M34929; AAA35248.1; -.
DR   PIR; A36467; A36467.
KW   Lysine biosynthesis; Oxidoreductase; NAD.
FT   ACT_SITE    205    205       BY SIMILARITY.
SQ   SEQUENCE   369 AA;  40612 MW;  B1B749FA008B8A36 CRC64;

  Query Match             3.3%;  Score 172;  DB 1;  Length 369;
  Best Local Similarity   20.5%;  Pred. No. 0.00063;
  Matches   81;  Conservative   64;  Mismatches  159;  Indels   92;  Gaps   12;

Qy     18 IIVQPSTRRIHHDAQYEDAGCEISEDLS-----ECGLIIGIKQPKLQMILSDRAYAFFSH 72
          : |: |  ||  | :: | |  : |: |     |  :|||:|:|:   :       :  |:|
Db     37 VFVEKSPLRIFDDQEFVDVGATLVEEGSWVSAPEDRMIIGLKELPEESFPLSHEHIQFAH 96

Qy     73 THKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRYL 132
           :| |     :| :      :|:| | :  |:|:|  |||  || ||
Db     97 CYKDQGGWKDVLSRFPAGNGTLYDLEFLEDDNGRRVAAFGFHAGFAG------------A 144

Qy    133 SLGYSTPFLSLGQSH----------MYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVF 182
          ::|  |   :  |:|           ||:       |  : :::|         | | |
Db    145 AIGVET--WAFQQTHPDSENLPGVSAYPN---ETELVDKIKKDLAAAVEKGSKLPTVLVI 199

Qy    183 TGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDIV 242
            :|    || :: : :         :||     |: :   :  :
Db    200 GALGRCGSGAIDLARKV--------GIPE----ENIIRWDMNETK--------------- 232

Qy    243 SHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLME 302
                    ||  :           : ||  | : :||:|  :  || :| | | :
Db    233 ----------KGGPF-----------QEIAD-ADIFINCIYLSQPIPPFINYDLLNKETR 270

Qy    303 TGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDILP 362
              :| |   | :    :     :|: : |    : :         |  :   ::| ||
Db    271 KLSVIVDVSADTTNPHNPVPVYTIATTFDHPTVPVETT--------AGPKLSVCSIDHLP 322

Qy    363 TEFSKEASQHFGNILSRLVASLASVKQPAELPSYLR 398
          :   :|||: |      |: ||  : |    | : |
Db    323 SLLPREASEAFS---EALLPSLLQLPQRDTAPVWTR 355

RESULT    3
LYS1_SCHPO
ID   LYS1_SCHPO      STANDARD;      PRT;   368 AA.
AC   Q09694; Q9UTC1;
DT   01-NOV-1995 (Rel. 32, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Saccharopine dehydrogenase [NAD+, L-lysine forming] (EC 1.5.1.7)
DE   (Lysine--2-oxoglutarate reductase) (SDH).
GN   LYS3 OR SPAC227.18 OR SPAC2F7.01.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Zimmermann W., Wambutt R., McDougall R.C., Rajandream M.A.,
RA   Barrell B.G.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 320-368 FROM N.A.
RC   STRAIN=972;
RA   Gentles S., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: N6-(L-1,3-dicarboxypropyl)-L-lysine + NAD(+) +
CC       H(2)O = L-lysine + 2-oxoglutarate + NADH.
CC   -!- PATHWAY: 8TH STEP IN LYSINE BIOSYNTHESIS.
CC   -!- SIMILARITY: TO OTHER FUNGAL SACCHAROPINE DEHYDROGENASE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as  long  as its content  is  in  no  way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL133156; CAB61467.1; -.
DR   EMBL; Z50142; CAA90488.1; -.
KW   Lysine biosynthesis; Oxidoreductase; NAD.
FT   ACT_SITE    205    205       BY SIMILARITY.
SQ   SEQUENCE   368 AA;  41392 MW;  789AB01DB171ED13 CRC64;

  Query Match             3.0%;  Score 157;  DB 1;  Length 368;
  Best Local Similarity   20.6%;  Pred. No. 0.0059;
  Matches   83;  Conservative   63;  Mismatches  156;  Indels  100;  Gaps   14;

Qy         17 RIIVQPSTRRIHHDAQYEDAGCEISEDLS-----ECGLIIGIKQ-PKLQMILSDRAYAFF 70
              :| :: |::|   | ::|  |  :  : |     :   |||:|:  |:         :  |
Db         35 QITIERSSQRAFKDKEFERLGFPMVPEGSWRHAPKDAYIIGLKELPENDNSPLKHTHIQF 94

Qy         71 SHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAG----LIDFLHG 126
              :| :| |:    :| :       |:| | :  |:|:|  |||  || ||     : : |
Db         95 AHCYKNQEGWREVLSRFPAGNGLLYDLEFLQDDNGRRVAAFGYHAGFAGSAISCLVWAHQ 154

Qy        127 LGQRYLSLGYSTPFLSLGQSHMYPSLAA---AKAAVIVVAEEI-ATFGLPSGICPIVFVF 182
              |               |  :  :|::      |: |  ||  ::       :   | : |
Db        155 L---------------LHPNKQFPAIRPFPNEKSLVRHVARQVRLALKKNNNQYPRILVI 199

Qy        183 TGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDIV 242
                :|    || ::          | |:   |    |      ||:
Db        200 GALGRCGTGACDL----------ASKIGIPFD-NILRWDINETKK--------------- 233

Qy        243 SHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLME 302
                  |  :  : |                     : |||:|     |    :: |
Db        234 --GGPFTEITESD---------------------IFVNCIYLSMPIPKFCTVESLN---V 267

Qy        303 TGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVV------CL 356
                  |  |||::||          :|:   |   |: :    |  :|  ||        :
Db        268 PNRKLRVVCDVSCD----------TTNPNNPIPIYNVNTTFDHPTVEVKGVTTPPPLEVI 317

Qy        357 AVDILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLR 398
              ::| |||   :|:|: |      |: || ::|     | ::|
Db        318 SIDHLPTLLPRESSEAFS---EALIPSLLALKDVDNAPVWVR 356


RESULT   4
LYS1_YEAST
ID   LYS1_YEAST     STANDARD;      PRT;   373 AA.
AC   P38998;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Saccharopine dehydrogenase [NAD+, L-lysine forming] (EC 1.5.1.7)
DE   (Lysine--2-oxoglutarate reductase) (SDH).
GN   LYS1 OR YIR034C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1278B;
RA   Feller A.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
```

```
RN   [3]
RP   SEQUENCE OF 203-212.
RX   MEDLINE=80198528; PubMed=6769500;
RA   Ogawa H., Hase T., Fujioka M.;
RT   "Amino acid sequence of a peptide containing an essential cysteine
RT   residue of yeast saccharopine dehydrogenase (L-lysine-forming).";
RL   Biochim. Biophys. Acta 623:225-228(1980).
CC   -!- CATALYTIC ACTIVITY: N6-(L-1,3-dicarboxypropyl)-L-lysine + NAD(+) +
CC       H(2)O = L-lysine + 2-oxoglutarate + NADH.
CC   -!- PATHWAY: 8TH STEP IN LYSINE BIOSYNTHESIS.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SIMILARITY: TO OTHER FUNGAL SACCHAROPINE DEHYDROGENASE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X77362; CAA54551.1; -.
DR   EMBL; Z38061; CAA86194.1; -.
DR   PIR; S41936; S41936.
DR   SGD; S0001473; LYS1.
KW   Lysine biosynthesis; Oxidoreductase; NAD.
FT   ACT_SITE    205    205
FT   CONFLICT    309    309       V -> A (IN REF. 2).
SQ   SEQUENCE   373 AA;  41492 MW;  6135911779D1DC41 CRC64;

  Query Match             2.9%;  Score 153.5;  DB 1;  Length 373;
  Best Local Similarity   21.2%;  Pred. No. 0.01;
  Matches   85;  Conservative   60;  Mismatches  158;  Indels   97;  Gaps   14;

Qy         17 RIIVQPSTRRIHHDAQYEDAGCEIS-----EDLSECGLIIGIKQ-PKLQMILSDRAYAFF 70
              :| |: | :   :  :|  ||  |      :      :|||:|: |:          : |
Db         35 KIYVEDSPQSTFNINEYRQAGAIIVPAGSWKTAPRDRIIIGLKEMPETDTFPLVHEHIQF 94

Qy         71 SHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQR 130
              :| :| |     :| : ::   :|:| | :  | |:|  ||| :|| ||       || |
Db         95 AHCYKDQAGWQNVLMRFIKGHGTLYDLEFLENDQGRRVAAFGFYAGFAGA-----ALGVR 149

Qy        131 YLSLGYSTPFLSLGQSHM----------YPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
                         :  |:|           ||:    || |  | ::           | |
Db        150 ---------DWAFKQTHSDDEDLPAVSPYPN---EKALVKDVTKDYKEALATGARKPTVL 197

Qy        181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRD 240
              :   :|    || ::   :         :|:     |: |                   |
Db.       198 IIGALGRCGSGAIDLLHKV--------GIPD----ANILKW-----------------D 227

Qy        241 IVSHKDPTR--QFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQ 298
              |   |: :|   ||:                   | | : :||:|  |   |  ||::|
Db        228 I---KETSRGGPFDE------------------IPQADIFINCIYLSKPIAPFTNMEKLN 266

Qy        299 QLMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAV 358
                      :| |   | :    |     :|   :|      :         |  :  :::
Db        267 NPNRRLRTVVDVSADTTNPHNPIPIYTVATVFNKPTVLVPTT--------VGPKLSVISI 318

Qy        359 DILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLR 398
              | ||:   :|||: | :    |: ||  : |    | ::|
Db        319 DHLPSLLPREASEFFSH---DLLPSLELLPQRKTAPVWVR 355


RESULT   5
LYS1_CANAL
ID   LYS1_CANAL     STANDARD;      PRT;   382 AA.
AC   P43065;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Saccharopine dehydrogenase [NAD+, L-lysine forming] (EC 1.5.1.7)
DE   (Lysine--2-oxoglutarate reductase) (SDH).
```

```
GN   LYS1.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95012722; PubMed=7927784;
RA   Garrad R.C., Schmidt T.M., Bhattacharjee J.K.;
RT   "Molecular and functional analysis of the LYS1 gene of Candida
RT   albicans.";
RL   Infect. Immun. 62:5027-5031(1994).
CC   -!- CATALYTIC ACTIVITY: N6-(L-1,3-dicarboxypropyl)-L-lysine + NAD(+) +
CC       H(2)O = L-lysine + 2-oxoglutarate + NADH.
CC   -!- PATHWAY: 8TH STEP IN LYSINE BIOSYNTHESIS.
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER FUNGAL SACCHAROPINE DEHYDROGENASE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U13233; AAA21362.1; -.
KW   Lysine biosynthesis; Oxidoreductase; NAD.
FT   ACT_SITE    217    217       BY SIMILARITY.
SQ   SEQUENCE   382 AA;  42394 MW;  A3620191DF04B88F CRC64;

  Query Match             2.7%;  Score 144.5;  DB 1;  Length 382;
  Best Local Similarity   21.2%;  Pred. No. 0.04;
  Matches   88;  Conservative   68;  Mismatches  161;  Indels   99;  Gaps   18;

Qy      18 IIVQPSTRRIHHDAQYEDAGCEISEDLS-----ECGLIIGIKQ-PKLQMILSDRAYAFFS 71
           | |: |::      :||  | :|  : |     :  :| |:|: |: :       :  |:
Db      38 IYVEESSQSTFDIKEYEAVGAKIVPEGSWKTAPKERIIFGLKELPENETFPLIHEHIQFA 97

Qy      72 HTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRA----GLIDF---- 123
           | :| |     :| :  :    |:| | :  | |:|  ||| :|| |    |::|:
Db      98 HCYKDQAGWQDVLKRFPQGNGILYDLEFLENDQGRRVAAFGFYAGFAGAAIGVLDWSFKQ 157

Qy     124 LHG--LGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFV 181
           |:|   | :    |   | ::      ||:       | :  |:  |     |  |   |
Db     158 LNGNTKGTKGEGEGGELPGVT-----PYPNENELIKDVKIELEKALTKN--GGQYPKCLV 210

Qy     182 FTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDI 241
              :|    || ::|| :         :|:     |::|   :
Db     211 IGALGRCGSGAIDLFKKI--------GIPD----DNIAKWDMA----------------- 241

Qy     242 VSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLM 301
                   :  ||            | | :   : : :||:|  |  || :| : |
Db     242 --------ETAKGG----------PFQEIVD--SDIFINCIYLSKPIPPFINKEILN--- 278

Qy     302 ETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRY-------DPSKNSYHDDMEGAGVV 354
                |  : |:: |          :|:   |   |       :|:     |  :|  :
Db     279 NENRKLTTIVDVSAD----------TTNPHNPIPVYEIATVFNEPTVEVKLD--KGPKLS 326

Qy     355 CLAVDILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACI--AHAGRL 408
             ::| ||:   :|||: |      |: ||  :      | ::|   :   |  ||
Db     327 VCSIDHLPSLLPREASEFFA---KDLMPSLLELPNRDTSPVWVRAKQLFDKHVARL 379


RESULT   6
MAP2_RAT
ID   MAP2_RAT         STANDARD;      PRT;  1861 AA.
AC   P15146;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Microtubule-associated protein 2 (MAP 2) (MAP2B) [Contains: MAP2C].
```

```
GN   MAP2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.
RC   STRAIN=WISTAR; TISSUE=Brain;
RX   MEDLINE=90251471; PubMed=2339070;
RA   Kindler S., Schwanke B., Schulz B., Garner C.C.;
RT   "Complete cDNA sequence encoding rat high and low molecular weight
RT   MAP2.";
RL   Nucleic Acids Res. 18:2822-2822(1990).
RN   [2]
RP   SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.
RC   STRAIN=WISTAR; TISSUE=Brain;
RX   MEDLINE=91060576; PubMed=2174050;
RA   Kindler S., Schulz B., Goedert M., Garner C.C.;
RT   "Molecular structure of microtubule-associated protein 2b and 2c from
RT   rat brain.";
RL   J. Biol. Chem. 265:19679-19684(1990).
RN   [3]
RP   SEQUENCE OF 1-151; 1515-1694 AND 1726-1861 FROM N.A.
RX   MEDLINE=90221819; PubMed=2326166;
RA   Doll T., Papandrikopoulou A., Matus A.;
RT   "Nucleotide and amino acid sequences of embryonic rat MAP2c.";
RL   Nucleic Acids Res. 18:361-361(1990).
RN   [4]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE=89365159; PubMed=2770869;
RA   Papandrikopoulou A., Doll T., Tucker R.P., Garner C.C., Matus A.;
RT   "Embryonic MAP2 lacks the cross-linking sidearm sequences and
RT   dendritic targeting signal of adult MAP2.";
RL   Nature 340:650-652(1989).
RN   [5]
RP   SEQUENCE OF 1695-1725 FROM N.A.
RX   MEDLINE=94110302; PubMed=8282767;
RA   Doll T., Meichsner M., Riederer B.M., Honegger P., Matus A.;
RT   "An isoform of microtubule-associated protein 2 (MAP2) containing
RT   four repeats of the tubulin-binding motif.";
RL   J. Cell Sci. 106:633-640(1993).
CC   -!- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY
CC       STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO
CC       SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.
CC   -!- ALTERNATIVE PRODUCTS: VARIOUS FORMS OF MAP2 ARE PRODUCED BY
CC       ALTERNATIVE SPLICING OF THE SAME GENE. MAP2C, THE LOW MOLECULAR
CC       FORM OF MAP2, LACKS THE CENTRAL DOMAIN OF MAP2A/B.
CC   -!- DEVELOPMENTAL STAGE: MAP2C IS EXPRESSED DURING EMBRYONIC BRAIN
CC       DEVELOPMENT AND UNTIL POSTANATAL DAY 10. MAP2B IS EXPRESSED
CC       THROUGHOUT BRAIN DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 3 OR 4 TAU/MAP REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X51842; CAA36135.1; -.
DR   EMBL; X17682; CAA35667.1; -.
DR   EMBL; X71487; CAA50588.1; -.
DR   PIR; S07887; S07887.
DR   PIR; S10003; S10003.
DR   PIR; A37981; A37981.
DR   InterPro; IPR001084; Tubulin-bind.
DR   Pfam; PF00418; tubulin-binding; 4.
DR   PROSITE; PS00229; TAU_MAP; 3.
KW   Microtubules; Repeat; Alternative splicing; Calmodulin-binding.
FT   DOMAIN     1454   1474       CALMODULIN-BINDING (POTENTIAL).
FT   REPEAT     1664   1694       TAU/MAP MOTIF.
FT   REPEAT     1695   1725       TAU/MAP MOTIF.
```

```
FT   REPEAT     1726   1756       TAU/MAP MOTIF.
FT   REPEAT     1757   1788       TAU/MAP MOTIF.
FT   VARSPLIC    152   1514       MISSING (IN ISOFORM MAP2C).
FT   VARSPLIC   1695   1725       MISSING (IN ISOFORM WITH 3 TAU/MAP
FT                                REPEATS).
SQ   SEQUENCE   1861 AA;  202409 MW;  42DCF116D21EF54E CRC64;

  Query Match             2.7%;  Score 141.5;  DB 1;  Length 1861;
  Best Local Similarity   20.9%;  Pred. No. 0.69;
  Matches  167;  Conservative   99;  Mismatches  267;  Indels  267;  Gaps   41;

Qy      252 DKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLMETGCPLVGVC 311
            |:|| |  |   ||   :         : |   | :       |: :|   ||     |
Db      743 DEGDDYLPPT--TPAVEK---------IPCFPIESK----EEEDKTEQAKVTGGQTTQV- 786

Qy      312 DITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDILPTEFSKEASQ 371
                            || | ||    |:|  | :     | | :| | ||       ::
Db      787 ---------------ETSSESPF----PAKEYYKN-----GTV-MAPD-LPEMLDLAGTR 820

Qy      372 HFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMRNTMIDLAPAKTN 431
                  ||| ||:::  : |   |    | :| :                   | |   :
Db      821 ------SRL-ASVSADAEVARRKSVPSEAVVAESS----------------TGLPPVADD 857

Qy      432 PLPDKKYSTLVSLSGHLFDKFLI---NEALDIIETAG--GSFHLVRCEVGQSTD------ 480
              | |  | |  :   :|:|: :   :   |    :|  |||:       : ||
Db      858 SQPVKPDSQLEDMGYCVFNKYTVPLPSPVQDSENLSGESGSFY-------EGTDDKVRRD 910

Qy      481 ---DMSYSELEVGA----DDTATLDKIIDSLTS-----LANEHGGDHDAGQEIELAL--- 525
               |:|  |::: |     |  | :|     :|     |: |   |  |  :::  |
Db      911 LATDLSLIEVKLAAAGRVKDEFTAEKEASPPSSADKSGLSREFDQDRKANDKLDTVLEKS 970

Qy      526 --------------KIG-KVNEYETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICT 570
                          ::| ||  :   || ::   | ||         | : |:| |::
Db      971 EEHVDSKEHAKESEEVGDKVELFGLGVTYEQTSAKELITTKETAPERAEKGLSSVPEVAE 1030

Qy      571 YGVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLL 630
                   |||   : |   :|| :  :|         : |:| | :
Db     1031 VETTT-KADQGLDVAA---KKDDQSPLD---------IKVSDFGQM-------------- 1063

Qy      631 PASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMK 690
             ||  :  ||  ||||        ||: ::  |:| ::               |  |
Db     1064 -ASGMSVDAGKTIELKFE------VDQQLTLSSEAPQET--------------DSFMG-- 1100

Qy      691 MIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETI 750
             |: :| : |   : |     :  |                                : :
Db     1101 -IESSHVKDGAKVSETEVKEKVAKP--------------------------------DLV 1127

Qy      751 H---VDGHNLYESA---KRLRLREL-PAFALEHLPNRNSLIYGDLY-----------GIS 792
            |   ||    |||:   : | :  | |    :      |||  ::             :|
Db     1128 HQEAVDKEESYESSGEHESLTMESLKPDEGKKETSPETSLIQDEVALKLSVEIPCPPPVS 1187

Qy      793 KEASTI-YRATXRYEGFSEIMVTLSKTGFFDAANHPLL-QDTSRPTYKGFLDELLNNIST 850
            :  |:|  :|  : |      : | |    :  : | :  | ::|  :  : |
Db     1188 EADSSIDEKAEVQME-----FIQLPKEESTETPDIPAIPSDVTQPQPEAVVSE---PAEV 1239

Qy      851 INTDLDIEASGGYD------DDL-IARLLKLG--------CCKNKEIAVKTVKTIK---- 891
               : :||| | ||      | | |  ||  |        |    :  |::| ||:
Db     1240 RGEEEEIEAEGEYDKLLFRSDTLQITDLLVPGSREEFVETCPGEHKGVVESVVTIEDDFI 1299

Qy      892 ----------FLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYP 941
                       || |         |  |    |  |:|  | |:|:  ||:  | : |
Db     1300 TVVQTTTDEGELGSH---------SVRFAAPVQPEEERRPYPHDEELEVLMAAEAQAEPK 1350

Qy      942 DGQP-----AEKHQATLLEF 956
            || |      || :    |:
Db     1351 DGSPDAPATPEKEEVPFSEY 1370


RESULT   7
DNAK_BACME
ID   DNAK_BACME     STANDARD;      PRT;   604 AA.
AC   P05646;
```

```
DT   01-NOV-1988 (Rel. 09, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa
DE   protein) (HSP70).
GN   DNAK.
OS   Bacillus megaterium.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1404;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87231083; PubMed=3035506;
RA   Sussman M.D., Setlow P.;
RT   "Nucleotide sequence of a Bacillus megaterium gene homologous to the
RT   dnaK gene of Escherichia coli.";
RL   Nucleic Acids Res. 15:3923-3923(1987).
CC   -!- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
CC   -!- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y00154; CAA68348.1; -.
DR   HSSP; P04475; 1DG4.
DR   InterPro; IPR001023; HSP70.
DR   Pfam; PF00012; HSP70; 1.
DR   PRINTS; PR00301; HEATSHOCK70.
DR   PROSITE; PS00297; HSP70_1; 1.
DR   PROSITE; PS00329; HSP70_2; 1.
DR   PROSITE; PS01036; HSP70_3; 1.
KW   Chaperone; ATP-binding; Heat shock.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   604 AA;  65119 MW;  29DF856DC5FAEC14 CRC64;

  Query Match             2.6%;  Score 139.5;  DB 1;  Length 604;
  Best Local Similarity   21.9%;  Pred. No. 0.17;
  Matches  143;  Conservative   81;  Mismatches  238;  Indels  191;  Gaps   30;

Qy      428 AKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQST-DDMS--- 483
            | |||      :|::|:  |:                 |: | |  |  | |  :||
Db       56 AITNP------NTIISVKRHM-----------------GTDHKVEAEGKQYTPQEMSAII 92

Qy      484 ------YSELEVGADDTATLDKIIDSLTSLAN--EHGGDHDAGQ----EIE--------- 522
                  |:|  :|   |    | : :: :  |  |     |||:    |:|
Db       93 LQHLKGYAEEYLGEPVT----KAVITVPAYFNDAERQATKDAGKIAGLEVERIINEPTAA 148

Qy      523 -LALKIGKVNEYETDVTIDKGG----PKILILGAGRVCRPAAEFLASYPDICTYGVDDHD 577
             ||  : | :| :| :  | ||      || || |       |  |:  |    | || |
Db      149 ALAYGLEKTDEDQTVLVYDLGGGTFDVSILELGDG-----VFEVRATAGD-NRLGGDDFD 202

Qy      578 ADQIHVIVASLYQKD-----------------AEETVDGIENTTATQLDVADIGSLSDLV 620
               |  :||   :::                 ||:    :   |:||: :  |   :
Db      203 QVIIDYLVAEFKKENGVDLSKDKMALQRLKDAAEKAKKDLSGVTSTQISLPFI--TAGEA 260

Qy      621 SQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLD 680
              : : :||  | |    ||:             |: :|: : || ||||
Db      261 GPLHLEVSLSRAKFDELSAGL-------------VERTMAPVRQALKDAG---------- 297

Qy      681 PGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNP 740
                  :|   :|:               ||     |  : :  :   :|      | ||
Db      298 ------LSASELDK-----------VILVGGSTRIPAVQDAIKKETGQDP----HKGVNP 336

Qy      741 AVYKFLGETIHVDGHNLYESAKRLRLRELPAFAL---------EHLPNRNSLIYGDLYGI 791
                 ||  |   |  |    | : | ::   :|           |  ||: |      :
Db      337 DEVVALGAAI--QGGVLTGDVKDVVLLDVTPLSLGIETMGGVFTKLIERNTTIPTSKSQV 394
```

```
Qy    792 SKEASTIYRATXRY--EGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNIS 849
             |:    |   :  :|   :       | |   : |       |  :|        :
Db    395 FSTAADSQTAVDIHVLQGERPMSADNKTLGRFQLTDIP-------PAPRG--------VP 439

Qy    850 TINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPF 909
           |    ||: :|  :     |   ||   ||| |: |:|:   |   |  :: |
Db    440 QIEVSFDIDKNGIVN----VRAKDLG--TNKEQAI-TIKSSTGLSDDEIDRMVKEAEENA 492

Qy    910 DVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEF-GKVENGRSTTAM 968
          |   || |                 ||||:     |     : ||  :  ||||    |  |
Db    493 DADKQRKE-----------------EVELRNEADQLVFTTEKTLKDLEGKVEEAEVTKAN 535

Qy    969 ALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSGIKLVEKVE 1021
               : |||      || ::    |:  : |:     ||:::  :|| |: :
Db    536 EAKDALKAAIE-----KNDLEE---IKAKKDELQ----EIVQALTVKLYEQAQ 576


RESULT   8
SPT6_YEAST
ID   SPT6_YEAST     STANDARD;      PRT;  1451 AA.
AC   P23615;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Transcription initiation protein SPT6.
GN   SPT6 OR SSN20 OR CRE2 OR YGR116W OR G6169.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90356011; PubMed=2201908;
RA   Swanson M.S., Carlson M., Winston F.;
RT   "SPT6, an essential gene that affects transcription in Saccharomyces
RT   cerevisiae, encodes a nuclear protein with an extremely acidic amino
RT   terminus.";
RL   Mol. Cell. Biol. 10:4935-4941(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061913; PubMed=8905931;
RA   Hansen M., Albers M., Backes U., Coblenz A., Leuther H., Neu R.,
RA   Schreer A., Schaefer B., Zimmermann M., Wolf K.;
RT   "The sequence of a 23.4 kb segment on the right arm of chromosome VII
RT   from Saccharomyces cerevisiae reveals CLB6, SPT6, RP28A and NUP57
RT   genes, a Ty3 element and 11 new open reading frames.";
RL   Yeast 12:1273-1277(1996).
RN   [3]
RP   SH2 DOMAIN.
RX   MEDLINE=94151820; PubMed=8108857;
RA   McLennan A.J., Shaw G.;
RT   "A yeast SH2 domain.";
RL   Trends Biochem. Sci. 18:464-465(1993).
CC   -!- FUNCTION: IMPORTANT FOR TRANSCRIPTION INITIATION. THE SPT4/5/6
CC       COMPLEX IS REQUIRED FOR TRANSCRIPTION INITIATION. IT MAY NORMALLY
CC       ACT TO REPRESS TRANSCRIPTION AT A VARIETY OF LOCI, AND ALSO PLAYS
CC       A ROLE IN CHROMATIN STRUCTURE OR ASSEMBLY.
CC   -!- SUBUNIT: COMPLEX OF SPT4, SPT5 AND SPT6.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   -!- SIMILARITY: TO YEAST SPT5 AND C.ELEGANS EMB-5.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M34391; AAA35086.1; -.
DR   EMBL; Z72899; CAA97124.1; -.
DR   EMBL; Z72902; CAA97127.1; -.
DR   PIR; A36468; A36468.
DR   SGD; S0003348; SPT6.
DR   InterPro; IPR000980; SH2.
DR   Pfam; PF00017; SH2; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS50001; SH2; 1.
KW   Transcription regulation; Nuclear protein; Phosphorylation;
KW   SH2 domain.
FT   DOMAIN        1    484       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN        8     12       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN       77     85       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      120    125       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN     1257   1354       SH2.
SQ   SEQUENCE   1451 AA;  168290 MW;  0BE9922A59BD0483 CRC64;

  Query Match             2.4%;  Score 125.5;  DB 1;  Length 1451;
  Best Local Similarity   20.0%;  Pred. No. 5.1;
  Matches  168;  Conservative  130;  Mismatches  289;  Indels  255;  Gaps   52;

Qy    317 IGGSIEFINKSTSIERPF-FRYDPSKNSYHDDMEGAGVVCLAVDI-----LPTEFSKEAS 370
          || :|:|| |  ::| || : |   : :|    |  | :    |:     |  ||    :
Db    374 IGNAIKFITKE-NLEVPFIYAY---RRNYISSREKDGFLLTEDDLWDIVSLDIEFHSLVN 429

Qy    371 QHFGNILSRLVASLASVKQPAELPSYLRR---ACIAHAGRLTPLYEYI----PRMRNTMI 423
          :   : : |  | |  :  |  :  | :    | ||    |  :|:|:        | |
Db    430 KK--DYVQRFYAEL-HIDDPI-VTEYFKNQNTASIAELNSLQDIYDYLEFKYANEINEMF 485

Qy    424 DLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQSTDDM- 482
               ||     ||:     |    ::||  :     :            ::| | :|:
Db    486 INHTGKTG----KKH-----LKNSSYEKFKASPLYQAVS-----------DIGISAEDVG 525

Qy    483 ----SYSELEVGADDTATLDKIIDSLTSLANEHGGD-----------HDAGQE---IELA 524
              |  ::    |  ::  | :: : |: | : ||            |  |:   :||:
Db    526 ENISSQHQIHPPVDHPSS--KPVEVIESILNANSGDLQVFTSNTKLAIDTVQKYYSLELS 583

Qy    525 LKIGKVNE---------YETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICTYGVDD 575
           |  |: |         |  || :   | | :  |            : | ||  | :
Db    584 -KNTKIREKVRSDFSKYYLADVVLTAKGKKEIQKG------------SLYEDI-KYAI-- 627

Qy    576 HDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLPASFH 635
           :   :|       ::| :  :  :|  :   | |        : || : :  |   :
Db    628 -NRTPMH------FRRDPDVFLKMVEAESLNLLSVK-----LHMSSQAQYIEHLFQIALE 675

Qy    636 AA-IAGVCIELK--KHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMI 692
              : : ||    : :     :|:   ::||  ||  :|  |:  :   : |   |  :
Db    676 TTNTSDIAIEWNNFRKLAFNQAMDKIFQDISQEVKD-NLTKNCQKLVAKTVRHKF-MTKL 733

Qy    693 DEA----HARKGKI-KAFTSYCGG--------------------------LPSPAAANPL 722
          |:|    : |  || |  :  ||                          : :|    ||
Db    734 DQAPFIPNVRDPKIPKILSLTCGQGRFGADAIIAVYVNRKGDFIRDYKIVDNPFDKTNPE 793

Qy    723 AYK-------FSWNPAGALRSGKNPAVYKF---LGETIH----VD--GHNL--------- 757
           ::        |  |     :| ||   ||   | | :|    ||  || :
Db    794 KFEDTLDNIIQSCQPNAIGINGPNPKTQKFYKRLQEVLHKKQIVDSRGHTIPIIYVEDEV 853

Qy    758 ---YESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGF-SEIMV 813
             |::::|         | :  ||:  |:    | |:  |  ::     |    || :
Db    854 AIRYQNSER---------AAQEFPNKPPLV---KYCIAL-ARYMHSPLLEYANLTSEEVR 900

Qy    814 TLSKTGFFDAANHP---LL--QDTSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLI 868
          :||         ||   ||  :  |      |:| ::| :|       :| :   |::
Db    901 SLS--------IHPHQNLLSSEQLSWALETAFVD-IVNLVS-------VEVNKATDNNYY 944

Qy    869 ARLLKLGCCKNKEIAVKTVKTIKFLG--------------LHEETQIPKGCSSPFDVICQ 914
          |  ||      |  |:  ::::: |               ||:   :    |: |  |
Db    945 ASALKYISGFGKRKAIDFLQSLQRLNEPLLARQQLITHNILHKTIFMN---SAGFLYISW 1001

Qy    915 RMEQRMAY---GHNEQDMVLLHHE----------VEVEY-PDGQPAEKHQATLLEFGKVE 960
            |:|  |    |:: |   :| |            :|| ||    :: | |: ||
Db   1002 N-EKRQKYEDLEHDQLDSTRIHPEDYHLATKVAADALEYDPDTIAEKEEQGTMSEF---- 1056
```

```
Qy        961 NGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSGIKLVEKV 1020
                                  : ||:     :  :  |  | |   ||  :::|:: :  :
Db       1057 --------------------IELLREDPDRRAKLESLNLESYAEELE--KNTGLRKLNNL 1094

Qy       1021 ET 1022
               |
Db       1095 NT 1096


RESULT   9
BACA_BACLI
ID   BACA_BACLI     STANDARD;      PRT;  5255 AA.
AC   O68006;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Bacitracin synthetase 1 (BA1) [Includes: ATP-dependent isoleucine
DE   adenylase (IleA) (Isoleucine activase); ATP-dependent cysteine
DE   adenylase (CysA) (Cysteine activase); ATP-dependent leucine adenylase
DE   (LeuA) (Leucine activase); ATP-dependent glutamate adenylase (GluA)
DE   (Glutamate activase); ATP-dependent isoleucine adenylase (IleA)
DE   (Isoleucine activase); Glutamate racemase (EC 5.1.1.3)].
GN   BACA.
OS   Bacillus licheniformis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1402;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 10716;
RX   MEDLINE=98089193; PubMed=9427658;
RA   Konz D., Klens A., Schoergendorfer K., Marahiel M.A.;
RT   "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC
RT   10716: molecular characterization of three multi-modular peptide
RT   synthetases.";
RL   Chem. Biol. 4:927-937(1997).
CC   -!- FUNCTION: ACTIVATES FIVE AMINO ACIDS, INCORPORATES TWO D-AMINO
CC       ACIDS, RELEASES AND CYCLIZES THE MATURE BACITRACIN.
CC   -!- CATALYTIC ACTIVITY: L-glutamate = D-glutamate.
CC   -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
CC       ANTIBIOTIC BACITRACIN.
CC   -!- SUBUNIT: LARGE MULTIENZYME COMPLEX OF BA1, BA2 AND BA3.
CC   -!- DOMAIN: CONSISTS OF FIVE MODULES AND ONE EPIMERIZATION DOMAIN IN
CC       THE FOURTH MODULE. EACH MODULE INCORPORATES ONE AMINO ACID INTO
CC       THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
CC       RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOLATION, CONDENSATION
CC       (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
CC       N METHYLATION (OPTIONAL).
CC   -!- MISCELLANEOUS: BACITRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
CC       DODECAPEPTIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
CC       ABUNDANT IS BACITRACIN A, A BRANCHED CYCLIC DODECAPEPTIDE. IT
CC       CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-
CC       GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION
CC       PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-
CC       HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
CC       TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
CC       CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
CC       PHE-9, AND ASP-11).
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF007865; AAC06346.1; -.
DR   HSSP; P14687; 1AMU.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; DUF4.
DR   InterPro; IPR003880; Phosphopant_attach.
DR   Pfam; PF00501; AMP-binding; 5.
DR   Pfam; PF00668; Condensation; 5.
DR   Pfam; PF00550; pp-binding; 5.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00455; AMP_BINDING; 5.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
KW   Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
KW   Multifunctional enzyme; Repeat.
FT   REPEAT       39    612       DOMAIN 1 (ISOLEUCINE-ACTIVATING).
FT   REPEAT     1109   1648       DOMAIN 2 (CYSTEINE-ACTIVATING).
FT   REPEAT     2124   2689       DOMAIN 3 (LEUCINE-ACTIVATING).
FT   REPEAT     3164   3732       DOMAIN 4 (GLUTAMINE-ACTIVATING).
FT   REPEAT     4668   5249       DOMAIN 5 (ISOLEUCINE-ACTIVATING).
FT   DOMAIN      621   1037       CYCLIZATION (POTENTIAL).
FT   DOMAIN      544    611       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1585   1652       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2621   2688       ACYL CARRIER (ACP) 3.
FT   DOMAIN     3664   3730       ACYL CARRIER (ACP) 4.
FT   DOMAIN     5171   5238       ACYL CARRIER (ACP) 5.
FT   BINDING     574    574       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1615   1615       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    2651   2651       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    3694   3694       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    5201   5201       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   5255 AA;  598254 MW;  906E8DD68450F85B CRC64;

  Query Match             2.3%;  Score 124;  DB 1;  Length 5255;
  Best Local Similarity   18.5%;  Pred. No. 46;
  Matches  141;  Conservative  119;  Mismatches  302;  Indels  202;  Gaps   37;

Qy         44 LSECGLIIGIKQPKLQMILSDRAYAFFSHT-------HKAQKENMPLLD-KILEERVSLF 95
              ||: |  :    |    :  |     | :|       |:: : :  ::| :::::
Db       2718 LSQTGSHVAYNMPFAMTLEGDFDIRRFENTLKNMVKRHESFRTSFVMIDGEVMQQIEKEI 2777

Qy         96 DYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRYLSLGYSTPFLSLGQSHMYPS-LAA 154
              |:::   | || |           :  |:              ||      |:  : |
Db       2778 DFQVAYSDIGKES-------AEEKIKSFIR-------------PF------HLEKAPLLR 2811

Qy        155 AKAAVIVVAEEIATFG----LPSGICPIVFVFTGVGNVSQG-AQEIFKLLPHTFVDAEKL 209
              |:   :   | :  |     :  |:   :|:   :| : :| : : | :    : : |
Db       2812 AEVVKLNEREHLLMFDMHHIISDGVSTDIFI-QELGALYEGKSLKPFHIQYKDYAEWE-- 2868

Qy        210 PEIFQARNLSKQSQSTKRVFQLYGCVVTSRDI------VSHKDPTRQFDKGDYYAHPEHY 263
                     |   :|:  ||  : |       ||      : || |  :  :||
Db       2869 -------NSHARSEELKRQ-EEYWLKTYKGDIPVLDLPIDHKRPLTKSSEGD-------- 2912

Qy        264 TPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLMETGCPLVGVCDITCDIGGSIEF 323
                             |      : |  | :| :     | |  :  :      : |
Db       2913 --------------TVTAAIESETFRKLQHMAK-----ENGVTMYMLL-----LAGYTAL 2948

Qy        324 INKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDILPTEFSKEASQHFGNILSRLVAS 383
              ::| |  :       |:    |:|::   :: : |: |      |  : | : |  :  :
Db       2949 LSKYTG-QEDIIVGTPAAGRNHEDIQ--HLIGMFVNTLAIRNHPEGKKTFRDYLQEVKEN 3005

Qy        384 LASVKQPAELP--SYLRRACIAHAGRLTPLYEYIPRMRNTMIDLAPAKTNPLPDKKYSTL 441
                   :  : |    : :  |       ||:: :    ||  |: | :   |     | |
Db       3006 TLQAYENQDYPFEELVEKVNIKRDMARNPLFDTMLVYHNT--DVKPFEAEGL----RSRL 3059

Qy        442 VSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQSTDDMSYSELEVGADDTATLDKII 501
              | :   :  ||      ||  ||  :   :| ||  ||   :   :|          ::
Db       3060 VEIKRGI-SKF------DITVTASEAADGLRLEVEYSTTLFNKERME----------RLS 3102

Qy        502 DSLTSLANEHGGDHDAGQEIELALKIGKVNEYETDVTIDKGGPKILILGAGRVCRPAAEF 561
              : | ||  |   ||                    |: |:    :| :|  |   |   :|
Db       3103 EHLISLL-EQAADH-------------------PDIAIN----QIDVLTKGERHRVLYDF 3138

Qy        562 LASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVD-----GIENTTATQLDVADIGS 615
```

```
             :    |             : :  |::| ||:| |      | |  :  :|:     |
Db    3139 NRTDGVFCK-----------EMTIPELFEKQAEKTPDHPAVAFGDETISYRELN-ERANS 3186

Qy     616 LSDLVSQ----VEVVISLLPASFHAAIAGVCIELKKH----MVTASYVDESMSNLSQAAK 667
           |:  : |     :|:  :|       | |:   ||       :  :|  |    :|   |
Db    3187 LAFTLRQKGVGPDVIAGILTERSIEMIVGIMGILKAGGAYLPIDPAYPQE---RISYIVK 3243

Qy     668 DAGVTILCEMG-LDPG---------------IDHLMSMKMIDEAHARKGKIKAFTSYCGG 711
           |: |::||  | :|||                ||: ::|     |  | :  |:  |  |
Db    3244 DSDVSVLCAAGDVDPGEAYTGDIIRIDQTGQNDHVENLK-----HDIKPQHLAYVIYTSG 3298

Qy     712 LPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGH 755
                :   |       :    |  | |  :|:      |:| |
Db    3299 -----STGKPKGVMIEHHSVNNLVHGLNERIYQ------HLDAH 3331


RESULT  10
GCN1_YEAST
ID   GCN1_YEAST     STANDARD;      PRT;  2672 AA.
AC   P33892;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Translational activator GCN1.
GN   GCN1 OR YGL195W OR G1318.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93268304; PubMed=8497269;
RA   Marton M.J., Crouch D., Hinnebusch A.G.;
RT   "GCN1, a translational activator of GCN4 in Saccharomyces cerevisiae,
RT   is required for phosphorylation of eukaryotic translation initiation
RT   factor 2 by protein kinase GCN2.";
RL   Mol. Cell. Biol. 13:3541-3556(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE=97197971; PubMed=9046087;
RA   Coglievina M., Klima R., Bertani I., Delneri D., Zaccaria P.,
RA   Bruschi C.V.;
RT   "Sequencing of a 40.5 kb fragment located on the left arm of
RT   chromosome VII from Saccharomyces cerevisiae.";
RL   Yeast 13:55-64(1997).
CC   -!- FUNCTION: TRANSLATIONAL ACTIVATOR OF GCN4. MAY BE INVOLVED IN
CC       SENSING CHARGED TRNA AND STIMULATING THE KINASE ACTIVITY OF GCN2
CC       IN AMINO ACID-STARVED CELLS. REQUIRED IN VIVO FOR THE
CC       PHOSPHORYLATION OF EIF-2-ALPHA ON SERINE-52 BY THE PROTEIN KINASE
CC       GCN2.
CC   -!- SUBUNIT: COMPONENT OF A HETEROMERIC COMPLEX THAT INCLUDES GCN1 AND
CC       GCN20.
CC   -!- SIMILARITY: STRONG, TO S.POMBE SPAC18G6.05C.
CC   -!- SIMILARITY: CONTAINS 20 HEAT REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L12467; AAA34635.1; -.
DR   EMBL; X91837; CAA62949.1; -.
DR   EMBL; Z72717; CAA96907.1; -.
DR   PIR; A48126; A48126.
DR   SGD; S0003163; GCN1.
DR   InterPro; IPR000357; HEAT_repeat.
DR   PROSITE; PS50077; HEAT_REPEAT; 4.
KW   Translation regulation; Activator; Repeat.
FT   REPEAT      932    970       HEAT 1.
```

```
FT   REPEAT     1030   1067       HEAT 2.
FT   REPEAT     1099   1138       HEAT 3.
FT   REPEAT     1243   1281       HEAT 4.
FT   REPEAT     1284   1321       HEAT 5.
FT   REPEAT     1405   1442       HEAT 6.
FT   REPEAT     1484   1521       HEAT 7.
FT   REPEAT     1523   1559       HEAT 8.
FT   REPEAT     1561   1598       HEAT 9.
FT   REPEAT     1603   1640       HEAT 10.
FT   REPEAT     1641   1679       HEAT 11.
FT   REPEAT     1721   1758       HEAT 12.
FT   REPEAT     1760   1796       HEAT 13.
FT   REPEAT     1862   1903       HEAT 14.
FT   REPEAT     1905   1942       HEAT 15.
FT   REPEAT     1947   1984       HEAT 16.
FT   REPEAT     1985   2024       HEAT 17.
FT   REPEAT     2097   2134       HEAT 18.
FT   REPEAT     2290   2328       HEAT 19.
FT   REPEAT     2347   2384       HEAT 20.
SQ   SEQUENCE   2672 AA;  296693 MW;  980FDD03753E9D1C CRC64;

  Query Match             2.3%;  Score 121;  DB 1;  Length 2672;
  Best Local Similarity   18.4%;  Pred. No. 25;
  Matches  201;  Conservative  156;  Mismatches  399;  Indels  334;  Gaps   48;

Qy      11 NGPRVNRIIVQPSTRRIH-HDAQYEDAGCEISEDLSECGLIIGIKQ--PKLQMILSD--- 64
           :|| :  :|:    | :|   |  :   |:|   :    :::  |   | || :: :
Db    1559 DGPSL-ALIIHIIHRGMHDRSANIKRKACKI---VGNMAILVDTKDLIPYLQQLIDEVEI 1614

Qy      65 ------------RAYAFFSHTHKAQKENMP-----LLDKILEERVSLFDYELIVGDDGKR 107
                        | |  :   :  :|  |     ||| : :|  |        ||
Db    1615 AMVDPVPNTRATAARALGALVERLGEEQFPDLIPRLLDTLSDESKS--------GD---- 1662

Qy     108 SLAFGKFAGRAGLIDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIA 167
                :      | : : |||   |               | |:: |           :
Db    1663 -----RLGSAQALAEVISGLGLTKL-------------DEMLPTILAG----------VT 1694

Qy     168 TFG--LPSGICPIVF--------------------VFTGVG----NVSQGAQEIFKLLPH 201
            |   :  |  |::                     :  :|:    |:   | :  ||:
Db    1695 NFRAYIREGFMPLLLFLPVCFGSQFAPYINQIIQPILSGLADNDENIRDTALKAGKLIVK 1754

Qy     202 TF----VDAEKLPEIFQAR-------NLSKQSQSTKRVFQLYGCVVTSRDIVSHKDPTR- 249
            :    ||   |||: :          ||    : : :||: |  ::||:  | :|
Db    1755 NYATKAVDL-LLPELERGMFDENDRIRLSSVQLTGELLFQVTG--ISSRNEFSEEDGDHN 1811

Qy     250 -QF----------DKGD------YYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLL 292
            :|          |: |      :    :    |    :  : ::: |     |   | |
Db    1812 GEFSGKLVDVLGQDRRDRILAALFVCRNDTSGIVRATTVDIWKALVPNTPRAVKEILPTL 1871

Qy     293 NMDQLQQLMETGCPLVGVC-----DITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDD 347
               :  |  :   |  :      |:   :||:      : |   |       : |  |
Db    1872 TGMIVTHLASSSNVLRNIAAQTLGDLVRRVGGN------ALSQLLPSLEESLIETSNSDS 1925

Qy     348 MEGAGVVCLAVDILPTEFSKEASQHFGN-----ILSRLVASLASVKQPAELPSYLRRACI 402
            :|   ||:|: |    | |    | :     | : |:   |:|:: | |   : :  :
Db    1926 RQG---VCIALYELIESASTETISQFQSTIVNIIRTALIDESATVREAAALSFDVFQDVV 1982

Qy     403 AHAGRLTPLYEYIPRMRNTM-----IDLAPAKTNPLPDKKYSTLV-----SLSGHLFDKF 452
                 | : | :| : : :      | |      :  ||   :      :|     | |
Db    1983 GK----TAVDEVLPYLLHMLESSDNSDFALLGLQEIMSKKSDVIFPILIPTLLAPPIDAF 2038

Qy     453 LINEALDIIETAGGSFH---------LVRCEVGQSTDDMSYSELEVGADDTATLDKIIDS 503
             :    : | || : :         ||   :| | |: :   ||:       ||::  |
Db    2039 RASALGSLAEVAGSALYKRLSIIINALVDAIIGTSEDESTKGALEL------ALDRVFLS 2092

Qy     504 LTSLANEHGGDHDAGQEIELALKIGKVNEYETDVTIDKGGPKILILGAGRVCRPAAEFLA 563
           :    |:  | |   |:|   ||   : :            :| :|          | |
Db    2093 V----NDDEGLHPLLQQIMSLLKSDNIEK------------RIAVL----------ERLP 2126

Qy     564 SYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQV 623
           :: |      | :  : :   : ||  :| :  |:|  |  :| |   |  :|  ||
Db    2127 NFFDKTVLDFDVYIPNFVSHAILSLDDED-QRVVNGNFNALSTLLKKVDKPTLEKLVKPA 2185
```

```
Qy    624 EVVISL--------------------LPASFHA--------------AIAGV-----CIE 644
          :  ::|                    ||   |               ||| |
Db   2186 KQSLALTGRQGQDVAAFKLPRGPNCVLPIFLHGLMYGSNDEREESALAIADVVSKTPAAN 2245

Qy    645 LKKHMVTAS-----YVDESMSNLSQAAKDAGVTIL---CEMGLDPGIDHLMS--MKMIDE 694
          ||  :   :      | |  |:  :||    : :|     | | | |  |    :| : :
Db   2246 LKPFVSVITGPLIRVVGERFSSDIKAAILFALNVLFIKIPMFLRPFIPQLQRTFVKSLSD 2305

Qy    695 AHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDG 754
          |     :::|     | |       :||  :       |   |   |: | | | |   |
Db   2306 ATNETLRLRA-AKALGALIEHQPRVDPLVIELVTGAKQATDEGVKTAMLKALLEVIMKAG 2364

Qy    755 HNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVT 814
            | |::|                |  :|:  :: | : : :  |         :::: :
Db   2365 SKLNENSK---------------TNIVNLVEEEMLGSNDKLAVAY---------AKLIGS 2400

Qy    815 LSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDIEA--------------- 859
          ||:    | | | :|||                   :| ||| |
Db   2401 LSEILSNDEA-HKILQD-----------------KVLNADLDGETGKFAILTLNSFLKDA 2442

Qy    860 -----SGGYDDDLIARLLKLGCCKNKEIAVK----------TVKTIKFLGLHEETQIP-- 902
               : |  |: :: :|          |::          |:   | | |  | : |
Db   2443 PTHIFNTGLIDEFVSYILN---------AIRSPDVYFGENGTIAAGKLLLLEGEKRSPFV 2493

Qy    903 -KGCSSPFDV 911
           |  : || :
Db   2494 KKDAAEPFKI 2503


RESULT  11
CETP_HUMAN
ID   CETP_HUMAN     STANDARD;      PRT;   493 AA.
AC   P11597; Q13987; Q13988;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cholesteryl ester transfer protein precursor (Lipid transfer protein
DE   I).
GN   CETP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=87258172; PubMed=3600759;
RA   Drayna D., Jarnagin A.S., McLean J., Henzel W., Kohr W., Fielding C.,
RA   Lawn R.;
RT   "Cloning and sequencing of human cholesteryl ester transfer protein
RT   cDNA.";
RL   Nature 327:632-634(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90241928; PubMed=2334701;
RA   Agellon L.B., Quinet E.M., Gillette T.G., Drayna D.T., Brown M.L.,
RA   Tall A.R.;
RT   "Organization of the human cholesteryl ester transfer protein gene.";
RL   Biochemistry 29:1372-1376(1990).
RN   [3]
RP   SEQUENCE OF 1-15 FROM N.A.
RX   MEDLINE=97112972; PubMed=8943225;
RA   Oliveira C.F.O., Chouinard R.A., Agellon L.B., Bruce C., Ma L.,
RA   Walsh A., Breslow J.L., Tall A.R.;
RT   "Human cholesteryl ester transfer protein gene proximal promoter
RT   contains dietary cholesterol positive responsive elements and mediates
RT   expression in small intestine and periphery while predominant liver
RT   and spleen expression is controlled by 5'-distal sequences. Cis-acting
RT   sequences mapped in transgenic mice.";
RL   J. Biol. Chem. 271:31831-31838(1996).
RN   [4]
RP   SEQUENCE OF 1-27 FROM N.A.
RX   MEDLINE=97473500; PubMed=9332354;
RA   Williams S., Hayes L., Elsenboss L., Williams A., Andre C.,
RA   Abramson R., Thompson J.F., Milos P.M.;
RT   "Sequencing of the cholesteryl ester transfer protein 5' regulatory
RT   region using artificial transposons.";
RL   Gene 197:101-107(1997).
RN   [5]
RP   SEQUENCE OF 9-493 FROM N.A.
RC   TISSUE=Liver;
RA   Dinchuk J.E., Hart J.T., Wirak D.O.;
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   VARIANT GLY-459.
RX   MEDLINE=94013514; PubMed=8408659;
RA   Takahashi K., Jiang X.-C., Sakai N., Yamashita S., Hirano K., Bujo H.,
RA   Yamazaki H., Kusunoki J., Miura T., Kussie P., Matsuzawa Y., Saito Y.,
RA   Tall A.;
RT   "A missense mutation in the cholesteryl ester transfer protein gene
RT   with possible dominant effects on plasma high density lipoproteins.";
RL   J. Clin. Invest. 92:2060-2064(1993).
CC   -!- FUNCTION: CETP IS AN EXTREMELY HYDROPHOBIC PROTEIN INVOLVED IN THE
CC       TRANSFER OF INSOLUBLE CHOLESTERYL ESTERS IN THE REVERSE TRANSPORT
CC       OF CHOLESTEROL.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED MAINLY IN THE LIVER.
CC   -!- DISEASE: PROBABLY INVOLVED IN THE DEVELOPMENT OF ATHEROSCLEROSIS.
CC   -!- SIMILARITY: BELONGS TO THE BPI/CETP/LBP/PLTP FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as  long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M30185; AAA51977.1; -.
DR   EMBL; M32998; AAA51978.1; -.
DR   EMBL; M32992; AAA51978.1; JOINED.
DR   EMBL; M32993; AAA51978.1; JOINED.
DR   EMBL; M32994; AAA51978.1; JOINED.
DR   EMBL; M32995; AAA51978.1; JOINED.
DR   EMBL; M32996; AAA51978.1; JOINED.
DR   EMBL; M32997; AAA51978.1; JOINED.
DR   EMBL; U71187; AAD14876.1; -.
DR   EMBL; AF027656; AAB86604.1; -.
DR   EMBL; M83573; AAB59388.1; -.
DR   PIR; A26941; A26941.
DR   MIM; 118470; -.
DR   InterPro; IPR001124; LBP_BPI_CETP.
DR   Pfam; PF01273; LBP_BPI_CETP; 1.
DR   Pfam; PF02886; LBP_BPI_CETP_C; 1.
DR   SMART; SM00328; BPI1; 1.
DR   SMART; SM00329; BPI2; 1.
DR   PROSITE; PS00400; LBP_BPI_CETP; 1.
KW   Lipid transport; Cholesterol metabolism; Glycoprotein; Signal;
KW   Atherosclerosis; Disease mutation.
FT   SIGNAL        1     17
FT   CHAIN        18    493       CHOLESTERYL ESTER TRANSFER PROTEIN.
FT   CARBOHYD    105    105       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    257    257       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    358    358       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    413    413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     459    459       D -> G (IN CETP DEFICIENCY).
FT                                /FTId=VAR_004172.
FT   CONFLICT    251    310       MISSING (IN REF. 5).
FT   CONFLICT    422    422       I -> V (IN REF. 2).
SQ   SEQUENCE   493 AA;  54770 MW;  16D47ACDC99B063C CRC64;

  Query Match             2.3%;  Score 120;  DB 1;  Length 493;
  Best Local Similarity   19.2%;  Pred. No. 2.3;
  Matches  115;  Conservative   75;  Mismatches  196;  Indels  212;  Gaps   30;
```

```
Qy     503 SLTSLANEH----GGDHDAGQEIELALKIGKVNEYETDVTIDKGGPKILILG--AGRVCR 556
           :|  | | |    |  |:||    :  :| |              | :|:|     :|  :
Db       8 TLALLGNAHACSKGTSHEAG----IVCRITK---------------PALLVLNHETAKVIQ 49

Qy     557 PAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSL 616
            | :  ||||||          ::  :::  :          |: |   : | :|
Db      50 TAFQ-RASYPDI--------TGEKAMMLLGQVKY--------GLHNIQISHLSIAS---- 88

Qy     617 SDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDA---GVTI 673
               ||||                        :| |  :| |: |:|   |     | |
Db      89 ----SQVE------------------------LVEAKSIDVSIQNVSVVFKGTLKYGYTT 120

Qy     674 LCEMGLDPG----IDHLMSMKMIDEAHARKGKIKAFTSYC-----------GGLPSPA-- 716
              :|:|      ||  : :::  :     |:::     |            |   |
Db     121 AWWLGIDQSIDFEIDSAIDLQINTQLTCDSGRVRTDAPDCYLSFHKLLLHLQGEREPGWI 180

Qy     717 ------------------------AANNPLAYKFSWNPAGALRSGKNPAVYKFLGETI- 750
                                  :| :|       |  |  |         |: :
Db     181 KQLFTNFISFTLKLVLKGQICKEINVISNIMADFVQTRAASILSDGDIGVDISLTGDPVI 240

Qy     751 -------HVDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATX 803
                  | || :|::       :||      ||  :  : ||             :
Db     241 TASYLESHHKGHFIYKNVS----EDLP------LPTFSPTLLGD-------------SRM 277

Qy     804 RYEGFSE-IMVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDI--EAS 860
            |  |||  :  :|:|  | |     |:       :|  |:         ||: :|  |
Db     278 LYFWFSERVFHSLAKVAFQDGR---LMLSLMGDEFKAVLETW-----GFNTNQEIFQEVV 329

Qy     861 GGYDD--DLIARLLKLG--CCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRM 916
           ||:     :    ||:    |:|| : | :   :|||                     |
Db     330 GGFPSQAQVTVHCLKMPKISCQNKGVVVNSSVMVKFL---------------------FPRP 370

Qy     917 EQRMAYGHN-EQDMVLLHHEVEVEYP---------DGQPAEKHQATLLEFGKVENGRSTT 966
           :|: :  :  |:|:|     |:  |          | |   |  : | |
Db     371 DQQHSVAYTFEEDIVT---TVQASYSKKKLFLSLLDFQITPKTVSNLTESSSESIQSFLQ 427

Qy     967 AMALTVGIPAAIGAL-LLLKNKVQTKGV-----IRPLQPEIYVPALEILESSGIKLVE 1018
           :|   ||||  :  | ::    : :|||     | |          ||:   |  |::
Db     428 SMITAVGIPEVMSRLEVVFTALMNSKGVSLFDIINP----------EIITRDGFLLLQ 475

RESULT  12
MAP2_MOUSE
ID   MAP2_MOUSE     STANDARD;      PRT;  1828 AA.
AC   P20357;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Microtubule-associated protein 2 (MAP 2).
GN   MAP2 OR MTAP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89083571; PubMed=3205744;
RA   Wang D., Lewis S.A., Cowan N.J.;
RT   "Complete sequence of a cDNA encoding mouse MAP2.";
RL   Nucleic Acids Res. 16:11369-11370(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89043973; PubMed=3142041;
RA   Lewis S.A., Wang D., Cowan N.J.;
RT   "Microtubule-associated protein MAP2 shares a microtubule binding
RT   motif with tau protein.";
RL   Science 242:936-939(1988).
CC   -!- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY
CC       STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO
CC       SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.
CC   -!- SIMILARITY: CONTAINS 3 TAU/MAP REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M21041; AAA39490.1; -.
DR   PIR; S06467; S06467.
DR   PIR; A40115; A40115.
DR   MGD; MGI:97175; Mtap2.
DR   InterPro; IPR001084; Tubulin-bind.
DR   Pfam; PF00418; tubulin-binding; 3.
DR   PROSITE; PS00229; TAU_MAP; 2.
KW   Microtubules; Repeat; Calmodulin-binding.
FT   DOMAIN     1452   1472       CALMODULIN-BINDING (POTENTIAL).
FT   REPEAT     1662   1692       TAU/MAP MOTIF.
FT   REPEAT     1693   1723       TAU/MAP MOTIF.
FT   REPEAT     1724   1755       TAU/MAP MOTIF.
SQ   SEQUENCE   1828 AA;  198980 MW;  200BC59E360538CA CRC64;

  Query Match              2.3%;  Score 119;  DB 1;  Length 1828;
  Best Local Similarity   20.4%;  Pred. No. 19;
  Matches  160;  Conservative  111;  Mismatches  280;  Indels  232;  Gaps   41;

Qy     252 DKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLMETGCPLVGVC 311
           |:|| |  |   ||   :         : |   | :       |: :|   ||
Db     741 DEGDDYLPPT--TPAVEK---------MPCFPIESK----EEEDKAEQAKVTG------- 778

Qy     312 DITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDILPTEFSKEASQ 371
                 | :|:     || | ||    |:|  | :     | | :| | ||
Db     779 ------GQTIQV---ETSSESPF----PAKEYYKN-----GTV-MAPD-LPEMLD----- 813

Qy     372 HFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMRNTMIDLAPAKTN 431
                |:   : ||||   ||:    ||  :     |      :| :         |  :
Db     814 -----LAGTRSRLASVSADAEV---ARRKSVPSEAMLAESSTSLPPV---------ADES 856

Qy     432 PLPDKKYSTLVSLSGHLFDKFLI---NEALDIIETAG--GSFHLVRCEVGQSTD------ 480
           |:  |  | |  :   :|:|: :   :   |    :|  |||:       : ||
Db     857 PVTVKPDSQLEDMGYCVFNKYTVPLPSPVQDSENLSGESGSFY-------EGTDDKVRRD 909

Qy     481 ---DMSYSELEVGA----DDTATLDKIIDSLTS-----LANEHGGDHDAGQEIELALKI- 527
              |:|  |::: |     | |  :|     ||     |: |   |  |  :::  |:
Db     910 LATDLSLIEVKLAAAGRVKDEFTAEKEATPPTSADKSGLSREFDHDRKANDKLDTVLEKS 969

Qy     528 -----------------GKVNEYETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICT 570
                            |||  :   :| |:   | ||           : |:| |::
Db     970 EEHIDSKEHAKESEEMGGKVELFGLGITYDQASTKELITTKDTSPEKTEKGLSSVPEVA- 1028

Qy     571 YGVDDHDADQIHVIVASLYQKDAEETV---DGIE----NTTATQLD--VADIGSLSDLVS 621
                                : | |     |::        :|||  |:| | :
Db    1029 ---------------------EVEPTTKADQGLDFAATKAEPSQLDIKVSDFGQM----- 1062

Qy     622 QVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDP 681
                     ||     ||  ||||        | : ::  |:| ::|       ||::
Db    1063 ----------ASGMNVDAGKAIELKFE------VAQELTLSSEAPQEAD----SFMGVES 1102

Qy     682 GIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPA 741
           |  |:     ::|   ::   |        |    | :   :|:          ||:: :
Db    1103 G--HIKEGGKVNETEVKEKVTKP------DLVHQEAVDKEESYE---------SSGEHES 1145

Qy     742 VYKFLGETIHVDGHNLYESAKRLRLRELPAFALE-HLPNRNSLIYGDLYGISKEASTIYR 800
           :     |::  |      | :   :::  |  |   :|    :   ||       ||  :
Db    1146 LTM---ESLKPDEGKKETSPETSLIQDEVALKLSVEIPCPPPVSEADL-------STDEK 1195

Qy     801 ATXRYEGFSEIMVTLSKTGFFDAANHPLL-QDTSRPTYKGFLDELLNNISTINTDLDIEA 859
              : |      : | |    :  : | :  | ::|  :  : |        : : :|||
Db    1196 GEVQME-----FIQLPKEESTETPDIPAIPSDVTQPQPEAIVSE---PAEVPSEEEEIEA 1247

Qy     860 SGGYDDDLI----ARLLKLGCCKNKEIAVKT-----------VKTIK--FLGLHEETQIP 902
            | ||  |      ::  |   :::|  |:|           | ||:  |: :  :|
Db    1248 GGEYDKLLFRSDTLQISDLLVSESREEFVETCPGELKGVVESVVTIEDDFITV-VQTTTD 1306
```

```
Qy     903 KGCSSP----FDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQP-----AEKHQATL 953
           :| |      |    |  |:|    |:|:  : :  | : |  || |      || :
Db    1307 EGESGSHSVRFAAPAQPEEERRPRPHDEELEIEMAAEAQAEPKDGSPDAPATPEKEEVAF 1366

Qy     954 LEF 956
            |:
Db    1367 SEY 1369


RESULT  13
ENO_ZYMMO
ID   ENO_ZYMMO      STANDARD;      PRT;   428 AA.
AC   P33675; O69010;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase) (2-phospho-D-
DE   glycerate hydro-lyase).
GN   ENO.
OS   Zymomonas mobilis.
OC   Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC   Zymomonas.
OX   NCBI_TaxID=542;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93015706; PubMed=1400207;
RA   Burnett M.E., Liu J., Conway T.;
RT   "Molecular characterization of the Zymomonas mobilis enolase (eno)
RT   gene.";
RL   J. Bacteriol. 174:6548-6553(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 31821 / ZM4 / CP4;
RA   Lee J., Jin S., Kang H.S.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: 2-phospho-D-glycerate = phosphoenolpyruvate +
CC       H(2)O.
CC   -!- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING
CC       THE DIMER.
CC   -!- PATHWAY: GLYCOLYSIS.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M99380; AAA27686.1; -.
DR   EMBL; AF086791; AAC70360.1; -.
DR   PIR; A45732; A45732.
DR   HSSP; P56252; 1PDZ.
DR   InterPro; IPR000941; Enolase.
DR   Pfam; PF00113; enolase; 1.
DR   PRINTS; PR00148; ENOLASE.
DR   ProDom; PD000902; Enolase; 1.
DR   PROSITE; PS00164; ENOLASE; 1.
KW   Lyase; Glycolysis; Magnesium.
FT   ACT_SITE    155    155       BY SIMILARITY.
FT   METAL       242    242       MAGNESIUM (BY SIMILARITY).
FT   METAL       287    287       MAGNESIUM (BY SIMILARITY).
FT   METAL       313    313       MAGNESIUM (BY SIMILARITY).
FT   CONFLICT    186    186       F -> S (IN REF. 2).
FT   CONFLICT    369    428       VETEDTTIADLAVATNCGQIKTGSLCRSERIAKYNQLMRIE
FT                                EELGSVAKYAGRSVLRKAK -> GGNRRHHDC (IN REF.
FT                                2).
SQ   SEQUENCE   428 AA;  45843 MW;  10943357EFFC2912 CRC64;
```

```
  Query Match              2.2%;  Score 118;  DB 1;  Length 428;
  Best Local Similarity   22.6%;  Pred. No. 2.5;
  Matches   81;  Conservative   50;  Mismatches  119;  Indels  108;  Gaps   20;

Qy     380 LVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMRNTMIDLAPAKTNPLPDKKYS 439
           |  |||  |  ||              |  ||| |:            |   |:|
Db     112 LGVSLAVAKAAAE--------------ARGLPLYRYVGG--------TAAHVLPVP---MM 147

Qy     440 TLVSLSGHL-----FDKFL---------INEALDIIETAGGS--FHLVRCEV---GQSTDD 481
            :|:   |      | :|:        ||||: |     |:  || :: |:   | :|:
Db     148 NIVNGGMHADNPIDFQEFMIAPVGASSINEAVRI-----GTEVFHTLKKELSAKGMNTNV 202

Qy     482 MSYSELEVGADD-TATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYETDVTID 540
                     |  :: || |:||::         :  |::: :||       |  |  |
Db     203 GDEGGFAPSLDSASSALDFIVDSISK------AGYKPGEDVFIALDAASSEFYNKDQNIY 256

Qy     541 KGGPKILILGAGRVCRPAAEFLASYPDIC----TYGVDD--HD-----------ADQIH 582
                  : | ||    :|: :  | ::|     | ::|   |             |::
Db     257 D------LKGEGRKL-TSAQLVDYYVELCGKYPIYSIEDLAEDDFEFWKILTEKLGDKVQ 309

Qy     583 VIVASLYQKDAEETVDGIENTTATQLDVA--DIGSLSDLVSQVEVV--ISLLPASFH--- 635
           ::   |:  : :   ||||   |  | |    |||||: :: | :    |  |   |
Db     310 LVGDDLFVTNVKRLSDGIERGIANSLLVKFNQIGSLSETLAAVNMANDASYTPVMSHRSV 369

Qy     636 ---------AAIAGVCIELKKHMVTAS----------YVDESMSNLSQAAKDAGVTIL 674
                     |:|  | ::|   :  |           ::|    |   || || ::|
Db     370 ETEDTTIADLAVATNCGQIKTGSLCRSERIAKYNQLMRIEE---ELGSVAKYAGRSVL 424


RESULT  14
RPO1_ASFB7
ID   RPO1_ASFB7     STANDARD;      PRT;  1450 AA.
AC   P42486;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   DNA-directed RNA polymerase subunit 1 (EC 2.7.7.6).
GN   NP1450L.
OS   African swine fever virus (strain BA71V) (ASFV).
OC   Viruses; dsDNA viruses, no RNA stage; Asfarviridae;
OC   African swine fever-like viruses.
OX   NCBI_TaxID=10498;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93281390; PubMed=8506138;
RA   Yanez R.J., Boursnell M., Nogal M.L., Yuste L., Vinuela E.;
RT   "African swine fever virus encodes two genes which share significant
RT   homology with the two largest subunits of DNA-dependent RNA
RT   polymerases.";
RL   Nucleic Acids Res. 21:2423-2427(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
RA   Rodriguez J.F., Vinuela E.;
RT   "Analysis of the complete nucleotide sequence of African swine fever
RT   virus.";
RL   Virology 208:249-278(1995).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SIMILARITY: TO THE CORRESPONDING SUBUNITS OF PROKARYOTIC AND
CC       EUKARYOTIC RNA POLYMERASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   -----------------------------------------------------------------------
DR   EMBL; Z21489; CAA79697.1; -.
DR   EMBL; U18466; AAA65328.1; -.
DR   InterPro; IPR000722; RNA_pol_A.
DR   InterPro; IPR002879; RNA_pol_A2.
DR   Pfam; PF00623; RNA_pol_A; 1.
DR   Pfam; PF01854; RNA_pol_A2; 1.
KW   Transferase; DNA-directed RNA polymerase; Transcription.
SQ   SEQUENCE   1450 AA;  163748 MW;  94D335C50B9A281B CRC64;

  Query Match              2.2%;  Score 116.5;  DB 1;  Length 1450;
  Best Local Similarity   18.0%;  Pred. No. 20;
  Matches  159;  Conservative  129;  Mismatches  306;  Indels  291;  Gaps   35;

Qy     82 PLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRYLSLGYSTPFL 141
          |::: |  |   ||           : :| |       :|  : |:||  ::     |
Db    718 PIMNSINPETNGLF-----------QMVATGAKGSNPNMIHIMAGIGQIEINTQRIQPQF 766

Qy    142 SLGQSHM-YPSLA-AAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVSQGAQEIFKLL 199
          | |:: : ||  |  |:|   :    ||      |:    |:|   | ::     | | |
Db    767 SFGRTLVYYPRFALEAQAYGFICNSYIA------GLTSPEFIF---GEMNGRFDLINKAL 817

Qy    200 PHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDIVSHKDPTRQFDKGDYYAH 259
            :                |    : | :| |  |:|                  |||
Db    818 STS----------------STGYANRKAIFGLQSCIV------------------DYY-- 841

Qy    260 PEHYTPVFHERIAPYASVIVNCMYWEK----------RFPPLLNMDQLQQLMETGCPLVG 309
                    |     : :|  :| |           ||  ::  ||
Db    842 ----------RRVSIDTRLVQQLYGEDGLDARQLETVRFETIMLSDQ------------- 878

Qy    310 VCDITCDIGGSIEFINKSTSIERPFF-----RYDPSKNSYHD---DMEGAGVVCLAVDIL 361
                     :|   | | |: | |     |    :: |     ::|      |  |:
Db    879 ----------ELEDKFKYTGIQSPLFEEEFSRLKKDRDKYRQIFLNVENFNFSQLLTDV- 927

Qy    362 PTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHA--------------GR 407
               ::   :  :|:  :: |  |   | :  | |::  :                  |
Db    928 -----RQVPVNVASIVKNILLSSTSGVLPFDEKSILQKYAMVKTFCKNLPYVFINNIQER 982

Qy    408 L-TPLYEYIPRMRNTM-----IDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLIN--EALD 459
          | ||:  |: |    |     |:||  ||  :  :: | :: |    : : |||  ||:
Db    983 LQTPIPVYLKRAAALMRMLIRIELATVKTLNITCEQMSAILDLIRLQYTQSLINYGEAVG 1042

Qy    460 IIETAGGSFHLVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLANEHG------- 512
          |:     |  | :                        ||    |:    |: |
Db   1043 ILAAQSVSEPLTQ----------------------YMLDSHHRSVAGGTNKSGIVRPQEI 1080

Qy    513 ---GDHDAGQEIELALKIGKVNEYETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDIC 569
                :| |  |: |:: |  | ||: |                    |: :|:  ::
Db   1081 FSAKPVEAEQSSEMLLRL-KNPEVETNKTY--------------------AQEIANSIELI 1120

Qy    570 TYGVDDHDADQIHVIVASL-----------YQKDAEETVDGIENTTATQ--LDVAD---- 612
          |:   :    | |::  :            :  | |   | :||    |   |:|:
Db   1121 TF---ERLILQWHLLYETYSSTKKNVMYPDFASDVEWMTDFLENHPLLQPPEDIANWCIR 1177

Qy    613 ---------IGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLS 663
                   : |:|     :| :|: | |              ||  |  |:  |: | :
Db   1178 LELNKTTMILKSIS-----LESIINSLRA--------------KHPNT--YIMHSVENTA 1216

Qy    664 QAAKDAGVTILCEMGLDP-----------GIDHLMSMKMIDEAHARKGKIKAFTSYCGGL 712
               :|: |:  : |               |  ::: ::|         |   |   |:
Db   1217 -----SGIPIIIRIYLRESAFRRSTNTRMATDEKIAVNVVD---------KLLNSTIRGI 1262

Qy    713 PSPAAANNPLAYKFSWNPAGAL----------RSGKN-------------PAVYKFLGET 749
          |    ||     :   :  | |           :| |               |   :|:|
Db   1263 PGIKNANVVKLMRHRVDAQGKLVRLDNIYAIKTNGTNIFGAMLDDNIDPYTIVSSSIGDT 1322

Qy    750 IHVDGHNLYESAKRLRLRELPAFALEHLPN-RNSLIYGDLYGISKEASTIYRATXRYEGF 808
          : : |    |:|::  : |:     :  || |: |:| ||   : : ::: :|
Db   1323 MELYG---IEAARQKIISEIRTVMGDKGPNHRHLLMYADLMTRTGQVTSLEKAGLNAREP 1379

Qy    809 SEIMVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNISTINT 853
          | ::: :: :          :     | |      |: ::  | |
```

```
Db   1380 SNVLLRMALSSPVQVLTDAAVDSAVNPIYGIAAPTLMGSVPRIGT 1424

RESULT 15
DYHC_MOUSE
ID   DYHC_MOUSE     STANDARD;      PRT;  4644 AA.
AC   Q9JHU4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Dynein heavy chain, cytosolic (DYHC) (Cytoplasmic dynein heavy chain).
GN   DNCHC1 OR DNCH1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB;
RA   Sasaki S., Shionoya A., Hirotsune S.;
RT   "Complete cDNA sequence of murine cytoplasmic dynein heavy chain.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC       MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC       ORGANELLES ALONG MICROTUBULES.
CC   -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC       INTERMEDIATE AND LIGHT CHAINS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AY004877; AAF91078.1; -.
DR   MGD; MGI:103147; Dnchc1.
KW   Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
FT   DOMAIN       48     69       COILED COIL (POTENTIAL).
FT   DOMAIN      179    200       COILED COIL (POTENTIAL).
FT   DOMAIN      453    476       COILED COIL (POTENTIAL).
FT   DOMAIN      541    564       COILED COIL (POTENTIAL).
FT   DOMAIN     1169   1201       COILED COIL (POTENTIAL).
FT   DOMAIN     1229   1250       COILED COIL (POTENTIAL).
FT   DOMAIN     1355   1371       COILED COIL (POTENTIAL).
FT   DOMAIN     2012   2040       MICROTUBULE-BINDING (POTENTIAL).
FT   DOMAIN     3187   3273       COILED COIL (POTENTIAL).
FT   DOMAIN     3394   3498       COILED COIL (POTENTIAL).
FT   DOMAIN     3735   3798       COILED COIL (POTENTIAL).
FT   NP_BIND    1904   1911       ATP (POTENTIAL).
FT   NP_BIND    2222   2229       ATP (POTENTIAL).
FT   NP_BIND    2593   2600       ATP (POTENTIAL).
FT   NP_BIND    2935   2942       ATP (POTENTIAL).
SQ   SEQUENCE   4644 AA;  532021 MW;  FE5B4E15DD479E1B CRC64;

  Query Match              2.2%;  Score 116.5;  DB 1;  Length 4644;
  Best Local Similarity   19.6%;  Pred. No. 1.2e+02;
  Matches  214;  Conservative  153;  Mismatches  406;  Indels  321;  Gaps   49;

Qy    114 FAGRAGLIDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATF---- 169
          :|||: | | |  |   : ||  : |   |    |  |     | |:  | :|  |
Db   2020 YAGRSNLPDNLKKL---FRSLAMTKPDRQLIAQVMLYSQGFRTAEVL--ANKIVPFFKLC 2074

Qy    170 --GLPS------GICPIVFVFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQ 221
             | |      |:  :  |    ||| :  : | |:        | : |   | || :|
Db   2075 DEQLSSQSHYDFGLRALKSVLVSAGNVKR--ERIQKIKREKEERGEAVDEGEIAENLPEQ 2132

Qy    222 SQSTKRVFQLYGCVVTSRDI------VSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYA 275
              : | :     : : ||      :|   |  |: :|:  |  |    |  |    |
```

```
Db   2133 EILIQSVCETMVPKLVAEDIPLLFSLLSDVFPGVQYHRGEMTALREELKKVCQEMYLTYG 2192

Qy    276 -SVIVNCMYWEKRFPPLLNMDQLQQLMETGCPLVGVCDITCDIGGSIEFINKSTSIERPF 334
              |  |: ||    :| : |: |:   |  :||        | |: :     ::||
Db   2193 DGEEVGGMWVEK----VLQLYQITQI-NHGLMMVG----PSGSGKSMAWRVLLKALER-- 2241

Qy    335 FRYDPSKNSYHDDMEGAGVVCLAVDILPTEFSKE-------------ASQHFGNILSRLV 381
                       :||   |   :|  |   ||:                 | ::| :::
Db   2242 -------------LEGVEGVAHIID--PKAISKDHLYGTLDPNTREWTDGLFTHVLRKII 2286

Qy    382 ASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMR-----NTMIDLAPAKTNPLP-- 434
           ::    |        :|  |   | : |  |::  :      | :: |   :   ||
Db   2287 DNVRGELQ--------KRQWIVFDGDVDP--EWVENLNSVLDDNKLLTLPNGERLSLPPN 2336

Qy    435 --------DKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQSTDDMSYSE 486
                  | ||:|| ::|                          || :   ::|:  ::
Db   2337 VRIMFEVQDLKYATLATVS--------------------------RCGMVWFSEDVLSTD 2370

Qy    487 LEVGADDTATLDKIIDSLTSLANEHG---------GDHDAGQE----------------- 520
          :         |:  :  | |:  : |         |  | |:|
Db   2371 M--------ILNNFLARLRSIPLDEGEDEAQRRRKGKEDEGEEAASPMLQIQRDAATIMQ 2422

Qy    521 ------------IELALKIGKVNEYETDVTIDKGGPKILILGA-----GRVCRPAAEFLA 563
                      :| | |:    |:  |:|      ::  ||:      : ||  |:: |
Db   2423 PYFTSNGLVTKALEHAFKL----EHIMDLT------RLRCLGSLFSMLHQACRNVAQYNA 2472

Qy    564 SYPDICTYGVDDHDADQIHVIVASLYQ-------KDAEETVDGIENTTATQLDVADIGSL 616
          ::||            | ::: | |:        |   |  : |   |   |  |    :
Db   2473 NHPDFPMQIEQLERYIQRYLVYAILWSLSGDSRLKMRAELGEYIRRITTVPLPTAPNVPI 2532

Qy    617 SDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGV----- 671
           |     || ||   : : | :    ||:: | | |   |  :  |     :| :
Db   2533 ID----YEVSISGEWSPWQAKVPQ--IEVETHKVAAP--DVVVPTLDTVRHEALLYTWLA 2584

Qy    672 ----TILC-------EMGLDPGIDHLMSMKMI----DEAHARKGKIKAFTSYC------- 709
               :||        | |   :  |  |:::      |   :  :| |  ||
Db   2585 EHKPLVLCGPPGSGKTMTLFSALRALPDMEVVGLNFSSATTPELLLKTFDHYCEYRRTPN 2644

Qy    710 GGLPSPA-----------AANNPLAYKF------------------------SW------ 728
          | : :|              | |   |:                        :|
Db   2645 GVVLAPVQLGKWLVLFCDEINLPDMDKYGTQRVISFIRQMVEHGAFYRTSDQTWVKLERI 2704

Qy    729 ------NPAGALRSGKNPAVYKFLGE--TIHVDG------HNLYESAKRLRLRELPAFAL 774
                ||      |: |  ::||     ::||         :| :  |  || :|:
Db   2705 QFVGACNP--PTDPGRKPLSHRFLRHVPVVYVDYPGPASLTQIYGTFNRAMLRLIPSLRT 2762

Qy    775 EHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAAN-------HP 827
             |   :::  : | :|:|  |  : |  :  :|   :|    | |:|
Db   2763 YAEPLTAAMV--EFYTMSQERFT--QDTQPHYIYSPREMTRWVRGIFEALRPLETLPVEG 2818

Qy    828 LLQDTSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTV 887
          |::  :    : | | |: :     || :|        |::|  ||     :|| |:
Db   2819 LIRIWAHEALRLFQDRLVEDEERRWTDENI--------DMVA--LKHFPNIDKEKAMS-- 2866

Qy    888 KTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVL------LHHEVEVEYP 941
          : | :     :  ||       | :  |::    :   | |: |      | | : ::
Db   2867 RPILYSNWLSKDYIPVDQEELRDYVKARLK---VFYEEELDVPLVLFNEVLDHVLRIDRI 2923

Qy    942 DGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEI 1001
            ||    |  ||  |    |::| :  :     | :  | : : ||  |        ::
Db   2924 FRQP----QGHLLLIGVSGAGKTTLSRFV-----AWMNGLSVYQIKVHRKYTGEDFDEDL 2974

Qy   1002 YVPALEILESSGIK 1015
                :|  || |
Db   2975 RT----VLRRSGCK 2984


Search completed: May 31, 2002, 15:51:00
Job time: 227 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          May 31, 2002, 15:47:43 ; Search time 41.27 Seconds
                                           (without alignments)
                                           4284.003 Million cell updates/sec

Title:           US-09-049-304A-122
Perfect score:   5286
Sequence:        1 CARLLLGGGKNGPRVNRIIV..........VPALEILESSGIKLVEKVET 1022

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters:      562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_19:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                    %
Result              Query
  No.     Score     Match Length DB  ID                 Description
-----------------------------------------------------------------------
    1     5281      99.9   1060  10  O04709              O04709 zea mays (m
    2     3430      64.9   1064  10  O04156              O04156 arabidopsis
    3     3426      64.8   1064  10  O04155              O04155 arabidopsis
    4     3411      64.5   1064  10  Q9SMZ4              Q9smz4 arabidopsis
    5     3381      64.0   1064  10  O04884              O04884 arabidopsis
    6   1670.5      31.6    465  10  Q947M5              Q947m5 arabidopsis
    7     1481      28.0    482  10  Q94BT4              Q94bt4 arabidopsis
    8   1466.5      27.7    972  5   Q9VLX0             Q9vlx0 drosophila
    9   1429.5      27.0    934  5   O44503             O44503 caenorhabdi
   10   1388.5      26.3    926  11   Q9Z1I9             Q9z1i9 mus musculu
   11   1384.5      26.2    926  11   Q99K67             Q99k67 mus musculu
   12   1363.5      25.8    926  4   Q9UDR5             Q9udr5 homo sapien
   13   1362.5      25.8    926  4   O95462             O95462 homo sapien
   14    746.5      14.1    448  3   Q9C2G6             Q9c2g6 neurospora
   15      717      13.6    450  3   Q9P4R4             Q9p4r4 magnaporthe
   16      716      13.5    450  3   O59711             O59711 schizosacch
   17    712.5      13.5    260  10  Q9FVF3              Q9fvf3 brassica ol
   18      709      13.4    449  3  Q96TW2              Q96tw2 penicillium
   19      608      11.5    177  10  O49176              O49176 brassica na
   20    602.5      11.4    748  3  Q96W51              Q96w51 cryptococcu
   21      583      11.0    252  10  Q9FVF4              Q9fvf4 brassica na
   22      385       7.3     91  10  Q9FUP1              Q9fup1 brassica na
   23      317       6.0    386  16  Q98IX7              Q98ix7 rhizobium l
   24    216.5       4.1    352  17  O59312              O59312 pyrococcus
   25      202       3.8    385  2  Q9AJC6              Q9ajc6 bacillus st
   26    192.5       3.6    373  17  Q978X1              Q978x1 thermoplasm
   27    176.5       3.3    373  17  Q9HKC0              Q9hkc0 thermoplasm
   28    145.5       2.8   1825  11  Q64715              Q64715 rattus norv
   29    145.5       2.8   1828  11  Q63724              Q63724 rattus ratt
   30    144.5       2.7    367  16  Q92LL0              Q92110 rhizobium m
   31    136.5       2.6    414  16  Q9KRL3              Q9krl3 vibrio chol
   32    136.5       2.6   1880  12  Q9DKW2              Q9dkw2 walrus cali
   33      135       2.6    405  16  Q9RUX8              Q9rux8 deinococcus
   34      134       2.5    367  16  Q98E01              Q98e01 rhizobium l
   35    133.5       2.5    846  2  Q9S6R3              Q9s6r3 escherichia
   36    133.5       2.5   1300  2  O32555              O32555 escherichia
   37    133.5       2.5   1300  2  O32591              O32591 escherichia
   38      133       2.5    792  10  Q9ARQ9              Q9arq9 oryza sativ
   39      132       2.5    398  16  Q55131              Q55131 synechocyst
   40    130.5       2.5   2832  2  Q93N05              Q93n05 agrobacteri
   41      130       2.5    407  17  O28990              O28990 archaeoglob
   42    127.5       2.4    411  17  O26953              O26953 methanother
   43    127.5       2.4    419  16  Q97RA6              Q97ra6 streptococc
   44      127       2.4    705  16  P74535              P74535 synechocyst
   45      127       2.4   1062  3  Q9P712              Q9p712 neurospora

                              ALIGNMENTS

RESULT   1
O04709
ID   O04709      PRELIMINARY;      PRT;  1060 AA.
AC   O04709;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LYSINE-KETOGLUTARATE REDUCTASE/SACCHAROPINE DEHYDROGENASE BIFUNCTIONAL
DE   ENZYME (LYSINE KETOGLUTARATE REDUCTASE/SACCHAROPINE
DE   DEHYDROGENASE).
GN   LKRSDH.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. AGROCERES F-352 COMMERCIAL HYBRID;
RA   Cord-Neto G., Kemper E.L., Arruda P.;
RT   "Lysine-ketoglutarate reductase/Saccharopine dehydrogenase
RT   bifunctional enzyme from maize seeds.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. B73;
RX   MEDLINE=20371120; PubMed=10908876;
RA   Arruda P., Kemper E.L., Papes F., Leite A.;
RT   "Regulation of lysine catabolism in higher plants.";
RL   Trends Plant Sci. 5:324-330(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. B73;
RA   Martinez-Moraes K.C., Kemper E.L., da Silva F.R., Vettore A.L.,
RA   Arruda P.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF003551; AAC18622.2; -.
DR   EMBL; AF271636; AAG21985.1; -.
SQ   SEQUENCE   1060 AA;  116302 MW;  00EB73105F1A4E68 CRC64;
```

```
Query Match             99.9%;  Score 5281;  DB 10;  Length 1060;
Best Local Similarity   99.8%;  Pred. No. 0;
Matches 1020;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     39 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 98

Qy     61 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 158

Qy    121 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 218

Qy    181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRD 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRD 278

Qy    241 IVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQL 300
          :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 MVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQL 338

Qy    301 METGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDI 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 METGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDI 398

Qy    361 LPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMRN 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    399 LPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMRN 458

Qy    421 TMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQSTD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 TMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQSTD 518

Qy    481 DMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYETDVTID 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 DMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYETDVTID 578

Qy    541 KGGPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGI 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 KGGPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGI 638

Qy    601 ENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMS 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    639 ENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMS 698

Qy    661 NLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANN 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    699 NLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANN 758

Qy    721 PLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNR 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    759 PLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNR 818

Qy    781 NSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGF 840
          |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    819 NSLIYGDLYGISKEASTIYRATLRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGF 878

Qy    841 LDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHEETQ 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    879 LDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHEETQ 938

Qy    901 IPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEFGKVE 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    939 IPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEFGKVE 998

Qy    961 NGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSGIKLVEKV 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    999 NGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSGIKLVEKV 1058

Qy   1021 ET 1022
          ||
Db   1059 ET 1060


RESULT   2
O04156
ID   O04156      PRELIMINARY;      PRT;  1064 AA.
AC   O04156;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LYSINE-KETOGLUTARATE REDUCTASE/SACCHAROPINE DEHYDROGENASE BIFUNCTIONAL
DE   ENZYME.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SSP. COLUMBIA;
RX   MEDLINE=98088001; PubMed=9426595;
RA   Epelbaum S., McDevitt R., Falco S.C.;
RT   "Lysine-ketoglutarate reductase and saccharopine dehydrogenase from
RT   Arabidopsis thaliana: nucleotide sequence and characterization.";
RL   Plant Mol. Biol. 35:735-748(1997).
DR   EMBL; U95759; AAB96826.1; -.
SQ   SEQUENCE   1064 AA;  117123 MW;  88AA67E332FD1ECB CRC64;

  Query Match             64.9%;  Score 3430;  DB 10;  Length 1064;
  Best Local Similarity   63.9%;  Pred. No. 2.4e-227;
  Matches  671;  Conservative  151;  Mismatches  176;  Indels   52;  Gaps   13;

Qy      1 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 60
          ||| || |||:   ::||:|||| :|||||| ||| |||||:|||:||||:|||||:|:|
Db     38 CAR-LLHGGKDRTGISRIVVQPSAKRIHHDALYEDVGCEISDDLSDCGLILGIKQPELEM 96

Qy     61 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 120
          || :|||||||||||||||||||||||| |||:| |||||||| ||| |||||:||||||
Db     97 ILPERAYAFFSHTHKAQKENMPLLDKILSERVTLCDYELIVGDHGKRLLAFGKYAGRAGL 156

Qy    121 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
          :||||||||||||||||||||||| |:|| |||||||||| | ||||: ||| ||||:||
Db    157 VDFLHGLGQRYLSLGYSTPFLSLGASYMYSSLAAAKAAVISVGEEIASQGLPLGICPLVF 216

Qy    181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIF-QARNLSKQSQSTKRVFQLYGCVVTSR 239
          |||| |||| ||||||||||||||:  ||||:| : : :|:   |||||:|:|||::||:
Db    217 VFTGTGNVSLGAQEIFKLLPHTFVEPSKLPELFVKDKGISQNGISTKRVYQVYGCIITSQ 276

Qy    240 DIVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQ 299
          |:| ||||:: ||| ||||||||| |||||:|:|| ||:||||||||||| ||:  |||
Db    277 DMVEHKDPSKSFDKADYYAHPEHYNPVFHEKISPYTSVLVNCMYWEKRFPCLLSTKQLQD 336

Qy    300 LMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVD 359
          | : | ||||:|||||||||||||:|::| |: ||||::|| |||:|||:| ||:|:|||
Db    337 LTKKGLPLVGICDITCDIGGSIEFVNRATLIDSPFFRFNPSNNSYYDDMDGDGVLCMAVD 396

Qy    360 ILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMR 419
          ||||||:||||||||:|||  | ||||: : ::||::|:||||:: | || |||||||||
Db    397 ILPTEFAKEASQHFGDILSGFVGSLASMTEISDLPAHLKRACISYRGELTSLYEYIPRMR 456

Qy    420 NTMIDLAPAKTNP------------LPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGS 467
                   |:||               : :: |||||||||||||||||||:|| ||||
Db    457 ---------KSNPEEAQDNIIANGVSSQRTFNILVSLSGHLFDKFLINEALDMIEAAGGS 507

Qy    468 FHLVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLAN---EHGGDHDAGQEIELA 524
          ||| :||:||| |  |||||||||||   ||:|||||| |||   ::   |    :|  :
Db    508 FHLAKCELGQSADAESYSELEVGADDKRVLDQIIDSLTRLANPNEDYISPHREANKI--S 565

Qy    525 LKIGKVNEYETDVTIDKGGPK------ILILGAGRVCRPAAEFLASYPDICT-------Y 571
```

```
              |||||| : | ::   |  |:      :|||||||||||||:||||   |  :       :
Db        566 LKIGKVQQ-ENEI---KEKPEMTKKSGVLILGAGRVCRPAADFLASVRTISSQQWYKTYF 621

Qy        572 GVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLP 631
              | |  :   :|||||||| |||:|||:|| :  | :|||:|  ||   ||||:||:||||
Db        622 GADSEEKTDVHVIVASLYLKDAKETVEGISDVEAVRLDVSDSESLLKYVSQVDVVLSLLP 681

Qy        632 ASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKM 691
              || || :|  |||||||:|||||||:  | | : || ||:||| ||||||||||:|:|||
Db        682 ASCHAVVAKTCIELKKHLVTASYVDDETSMLHEKAKSAGITILGEMGLDPGIDHMMAMKM 741

Qy        692 IDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIH 751
              |::|| :|||:|:|||||||||||||||||||||||||||||:|:|:||| ||  |: ||
Db        742 INDAHIKKGKVKSFTSYCGGLPSPAAANNPLAYKFSWNPAGAIRAGQNPAKYKSNGDIIH 801

Qy        752 VDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEI 811
              ||| |||:|| | |:  ||||||| ||||:||:||: |||  ||:||:| | |||||| |
Db        802 VDGKNLYDSAARFRVPNLPAFALECLPNRDSLVYGEHYGIESEATTIFRGTLRYEGFSMI 861

Qy        812 MVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIARL 871
              | |||| ||||:  : :|    | |:   |  :||     : | : |   | ::::  |:
Db        862 MATLSKLGFFDSEANQVLSTGKRITFGALLSNILNK----DADNESEPLAG-EEEISKRI 916

Qy        872 LKLGCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVL 931
              :|||   :|| | |  ||| ||| :|| ::|  | | ||  |  ||:::||  |||||||
Db        917 IKLG--HSKETAAKAAKTIVFLGFNEEREVPSLCKSVFDATCYLMEEKLAYSGNEQDMVL 974

Qy        932 LHHEVEVEYPDGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTK 991
              ||||||||: : :  ||| ||||||| ::||::||||| |||||||||||||:::|::|:
Db        975 LHHEVEVEFLESKRIEKHTATLLEFGDIKNGQTTTAMAKTVGIPAAIGALLLIEDKIKTR 1034

Qy        992 GVIRPLQPEIYVPALEILESSGIKLVEKVE 1021
              ||:|||: |:|:|||:||:: ||||:|| |
Db       1035 GVLRPLEAEVYLPALDILQAYGIKLMEKAE 1064

*
RESULT    3
O04155
ID   O04155      PRELIMINARY;      PRT;  1064 AA.
AC   O04155;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LYSINE-KETOGLUTARATE REDUCTASE/SACCHAROPINE DEHYDROGENASE BIFUNCTIONAL
DE   ENZYME.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LANDSBERG ERECTA;
RX   MEDLINE=98088001; PubMed=9426595;
RA   Epelbaum S., McDevitt R., Falco S.C.;
```

Inventors (not > 1 yr. from priority of SEQID 170 + 122)

```
RT   "Lysine-ketoglutarate reductase and saccharopine dehydrogenase from
RT   Arabidopsis thaliana: nucleotide sequence and characterization.";
RL   Plant Mol. Biol. 35:735-748(1997).
DR   EMBL; U95758; AAB96825.1; -.
SQ   SEQUENCE   1064 AA;  117157 MW;  820ACDE332FD1ECB CRC64;

  Query Match             64.8%;  Score 3426;  DB 10;  Length 1064;
  Best Local Similarity   63.8%;  Pred. No. 4.5e-227;
  Matches  670;  Conservative  151;  Mismatches  177;  Indels   52;  Gaps   13;

Qy          1 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 60
              ||| || |||:   ::||:|||| :|||||| ||| |||||:|||:||||:|||||:|:|
Db         38 CAR-LLHGGKDRTGISRIVVQPSAKRIHHDALYEDVGCEISDDLSDCGLILGIKQPELEM 96

Qy         61 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 120
              || :|||||||||||||||||||||||| |||:| |||||||| ||| |||||:||||||
Db         97 ILPERAYAFFSHTHKAQKENMPLLDKILSERVTLCDYELIVGDHGKRLLAFGKYAGRAGL 156

Qy        121 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
              :||||||||||||||||||||||| |:|| |||||||||| | ||||: ||| ||||:||
Db        157 VDFLHGLGQRYLSLGYSTPFLSLGASYMYSSLAAAKAAVISVGEEIASQGLPLGICPLVF 216

Qy        181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIF-QARNLSKQSQSTKRVFQLYGCVVTSR 239
              |||| |||| ||||||||||||||:  ||||:| : : :|:   |||||:|:|||::||:
Db        217 VFTGTGNVSLGAQEIFKLLPHTFVEPSKLPELFVKDKGISQNGISTKRVYQVYGCIITSQ 276

Qy        240 DIVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQ 299
              |:| ||||:: ||| ||||||||| |||||:|:|| ||:||||||||||| ||:  |||
Db        277 DMVEHKDPSKSFDKADYYAHPEHYNPVFHEKISPYTSVLVNCMYWEKRFPCLLSTKQLQD 336

Qy        300 LMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVD 359
              | : | ||||:|||||||||||||:|::| |: ||||::|| |||:|||:| ||:|:|||
Db        337 LTKKGLPLVGICDITCDIGGSIEFVNRATLIDSPFFRFNPSNNSYYDDMDGDGVLCMAVD 396

Qy        360 ILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMR 419
              ||||||:||||||||:|||  | ||||: : ::||::|:||||:: | || |||||||||
Db        397 ILPTEFAKEASQHFGDILSGFVGSLASMTEISDLPAHLKRACISYRGELTSLYEYIPRMR 456

Qy        420 NTMIDLAPAKTNP------------LPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGS 467
                       |:||               : :: |||||||||||||||||||:|| ||||
Db        457 ---------KSNPEEAQDNIIANGVSSQRTFNILVSLSGHLFDKFLINEALDMIEAAGGS 507

Qy        468 FHLVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLAN---EHGGDHDAGQEIELA 524
              ||| :||:||| |  |||||||||||   ||:|||||| |||   ::   |    :|  :
Db        508 FHLAKCELGQSADAESYSELEVGADDKRVLDQIIDSLTRLANPNEDYISPHREANKI--S 565

Qy        525 LKIGKVNEYETDVTIDKGGPK------ILILGAGRVCRPAAEFLASYPDICT-------Y 571
              |||||| : | ::   |  |:      :|||||||||||||:||||   |  :       :
Db        566 LKIGKVQQ-ENEI---KEKPEMTKKSGVLILGAGRVCRPAADFLASVRTISSQQWYKTYF 621

Qy        572 GVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLP 631
              | |  :   :|||||||| |||:|||:|| :  | :|||:|  ||   ||||:||:||||
Db        622 GADSEEKTDVHVIVASLYLKDAKETVEGISDVEAVRLDVSDSESLLKYVSQVDVVLSLLP 681

Qy        632 ASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKM 691
              || || :|  |||||||:|||||||:  | | : || ||:||| ||||||||||:|:|||
Db        682 ASCHAVVAKTCIELKKHLVTASYVDDETSMLHEKAKSAGITILGEMGLDPGIDHMMAMKM 741

Qy        692 IDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIH 751
              |::|| :|||:|:|||||||||||||||||||||||||||||:|:|:||| ||  |: ||
Db        742 INDAHIKKGKVKSFTSYCGGLPSPAAANNPLAYKFSWNPAGAIRAGQNPAKYKSNGDIIH 801

Qy        752 VDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEI 811
              ||| |||:|| | |:  ||||||| ||||:||:||: |||  ||:||:| | |||||| |
Db        802 VDGKNLYDSAARFRVPNLPAFALECLPNRDSLVYGEHYGIESEATTIFRGTLRYEGFSMI 861

Qy        812 MVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIARL 871
              | |||| ||||:  : :|    | |:   |  :||     : | : |   | ::::  |:
Db        862 MATLSKLGFFDSEANQVLSTGKRITFGALLSNILNK----DADNESEPLAG-EEEISKRI 916

Qy        872 LKLGCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVL 931
              :|||   :|| | |  ||| ||| :|| ::|  | | ||  |  ||:::||  |||||||
Db        917 IKLG--HSKETAAKAAKTIVFLGFNEEREVPSLCKSVFDATCYLMEEKLAYSGNEQDMVL 974

Qy        932 LHHEVEVEYPDGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTK 991
              ||||||||: : :  ||| ||||||| ::||::||||| |||||||||||||:::|::|:
Db        975 LHHEVEVEFLESKRIEKHTATLLEFGDIKNGQTTTAMAKTVGIPAAIGALLLIEDKIKTR 1034

Qy        992 GVIRPLQPEIYVPALEILESSGIKLVEKVE 1021
              ||:|||: |:|:|||:||:: ||||:|| |
Db       1035 GVLRPFEAEVYLPALDILQAYGIKLMEKAE 1064

RESULT    4
Q9SMZ4
ID   Q9SMZ4      PRELIMINARY;      PRT;  1064 AA.
AC   Q9SMZ4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
```

```
DE   LYSINE-KETOGLUTARATE REDUCTASE/SACCHAROPINE (EC 1.5.1.8).
GN   F4I10.80 OR AT4G33150.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Peters S.A., van Staveren M., Dirkse W., Stiekema W.,
RA   Hoheisel J., Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Peters S.A., van Staveren M., Dirkse W., Stiekema W., Mewes H.W.,
RA   Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035525; CAB36789.1; -.
DR   EMBL; AL161583; CAB80032.1; -.
KW   Oxidoreductase.
SQ   SEQUENCE   1064 AA;  117148 MW;  6CA4EBDB22898C7E CRC64;

  Query Match             64.5%;  Score 3411;  DB 10;  Length 1064;
  Best Local Similarity   63.6%;  Pred. No. 4.8e-226;
  Matches  668;  Conservative  151;  Mismatches  179;  Indels   52;  Gaps   13;

Qy          1 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 60
              ||| || |||:   ::||:|||| :|||||| ||| |||||:|||:||||:|||||:|:|
Db         38 CAR-LLHGGKDRTGISRIVVQPSAKRIHHDALYEDVGCEISDDLSDCGLILGIKQPELEM 96

Qy         61 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 120
              || :|||||||||||||||||||||||| |||:| |||||||| ||| |||||:||||||
Db         97 ILPERAYAFFSHTHKAQKENMPLLDKILSERVTLCDYELIVGDHGKRLLAFGKYAGRAGL 156

Qy        121 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
              :||||||||| | ||||||||||| |:|| |||||||||| | ||||: ||| ||||:||
Db        157 VDFLHGLGQRKLILGYSTPFLSLGASYMYSSLAAAKAAVISVGEEIASQGLPLGICPLVF 216

Qy        181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIF-QARNLSKQSQSTKRVFQLYGCVVTSR 239
              |||| |||| ||||||||||||||:  ||||:| : : :|:   |||||:|:|||::||:
Db        217 VFTGTGNVSLGAQEIFKLLPHTFVEPSKLPELFVKDKGISQNGISTKRVYQVYGCIITSQ 276

Qy        240 DIVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQ 299
              |:| ||||:: ||| ||||||||| |||||:|:|| ||:||||||||||| ||:  |||
Db        277 DMVEHKDPSKSFDKADYYAHPEHYNPVFHEKISPYTSVLVNCMYWEKRFPCLLSTKQLQD 336

Qy        300 LMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVD 359
              | : | ||||:|||||||||||||:|::| |: ||||::|| |||:|||:| ||:|:|||
Db        337 LTKKGLPLVGICDITCDIGGSIEFVNRATLIDSPFFRFNPSNNSYYDDMDGDGVLCMAVD 396

Qy        360 ILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMR 419
              ||||||:||||||||:|||  | ||||: : ::||::|:||||:: | || |||||||||
Db        397 ILPTEFAKEASQHFGDILSGFVGSLASMTEISDLPAHLKRACISYRGELTSLYEYIPRMR 456

Qy        420 NTMIDLAPAKTNP------------LPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGS 467
                       |:||               : :: |||||||||||||||||||:|| ||||
Db        457 ---------KSNPEEAQDNIIANGVSSQRTFNILVSLSGHLFDKFLINEALDMIEAAGGS 507

Qy        468 FHLVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLAN---EHGGDHDAGQEIELA 524
              ||| :||:||| |  |||||||||||   ||:|||||| |||   ::   |    :|  :
Db        508 FHLAKCELGQSADAESYSELEVGADDKRVLDQIIDSLTRLANPNEDYISPHREANKI--S 565

Qy        525 LKIGKVNEYETDVTIDKGGPK------ILILGAGRVCRPAAEFLASYPDICT-------Y 571
              |||||| : | ::   |  |:      :|||||||||||||:||||   | :       :
Db        566 LKIGKVQQ-ENEI---KEKPEMTKKSGVLILGAGRVCRPAADFLASVRTISSQQWYKTYF 621

Qy        572 GVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLP 631
              | |  :   :|||||||| |||:|||:|| :  | :|||:|  ||   ||||:||:||||
Db        622 GADSEEKTDVHVIVASLYLKDAKETVEGISDVEAVRLDVSDSESLLKYVSQVDVVLSLLP 681

Qy        632 ASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKM 691
              || || :|  |||||||:|||||||:  | | : || ||:||| ||||||||||:|:|||
Db        682 ASCHAVVAKTCIELKKHLVTASYVDDETSMLHEKAKSAGITILGEMGLDPGIDHMMAMKM 741

Qy        692 IDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIH 751
              |::|| :|||:|:|||||||||||||||||||||||||||||:|:|:||| ||  |: ||
Db        742 INDAHIKKGKVKSFTSYCGGLPSPAAANNPLAYKFSWNPAGAIRAGQNPAKYKSNGDIIH 801

Qy        752 VDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEI 811
              ||| |||:|| | |:  |||||||  |||:||:||: |||  ||:||:| | |||||| |
Db        802 VDGKNLYDSAARFRVPNLPAFALECFPNRDSLVYGEHYGIESEATTIFRGTLRYEGFSMI 861

Qy        812 MVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIARL 871
              | |||| ||||:  : :|    | |:   |  :||     : | : |   | ::::  |:
Db        862 MATLSKLGFFDSEANQVLSTGKRITFGALLSNILNK----DADNESEPLAG-EEEISKRI 916

Qy        872 LKLGCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVL 931
              :|||   :|| | |  ||| ||| :|| ::|  | | ||  |  ||:::||  |||||||
Db        917 IKLG--HSKETAAKAAKTIVFLGFNEEREVPSLCKSVFDATCYLMEEKLAYSGNEQDMVL 974

Qy        932 LHHEVEVEYPDGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTK 991
              ||||||||: : :  ||| ||||||| ::||::||||| |||||||||||||:::|::|:
Db        975 LHHEVEVEFLESKRIEKHTATLLEFGDIKNGQTTTAMAKTVGIPAAIGALLLIEDKIKTR 1034

Qy        992 GVIRPLQPEIYVPALEILESSGIKLVEKVE 1021
              ||:|||: |:|:|||:||:: ||||:|| |
Db       1035 GVLRPLEAEVYLPALDILQAYGIKLMEKAE 1064


RESULT    5
O04884
ID   O04884       PRELIMINARY;   PRT;  1064 AA.
AC   O04884;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LYSINE-KETOGLUTARATE REDUCTASE/SACCHAROPINE DEHYDROGENASE (EC 1.5.1.8)
DE   (SACCHAROPINE DEHYDROGENASE (NADP+, L-LYSINE FORMING)) (EC 1.5.1.9).
GN   LKR/SDH.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA; TISSUE=SEEDLING HYPOCOTYLS;
RA   Tang G., Miron D., Zhu-Shimoni J.X., Galili G.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: N6-(L-1,3-DICARBOXYPROPYL)-L-LYSINE + NADP(+)
CC       + H(2)O = L-LYSINE + 2-OXOGLUTARATE + NADPH.
DR   EMBL; U90522; AAB53975.1; -.
DR   EMBL; U90523; AAD00700.1; -.
KW   Oxidoreductase.
SQ   SEQUENCE   1064 AA;  117023 MW;  3DD075F9520ED7A9 CRC64;

  Query Match             64.0%;  Score 3381;  DB 10;  Length 1064;
  Best Local Similarity   63.3%;  Pred. No. 5.6e-224;
  Matches  665;  Conservative  150;  Mismatches  183;  Indels   52;  Gaps   13;

Qy          1 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 60
              ||| || |||:   ::||:|||| :|||||| ||| |||||:|||:||||:|||||:|:|
Db         38 CAR-LLHGGKDRTGISRIVVQPSAKRIHHDALYEDVGCEISDDLSDCGLILGIKQPELEM 96

Qy         61 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 120
              || :|||||||||||||||||||||||| |||:| |||||||| ||| |||||:||||||
```

```
Db         97 ILPERAYAFFSHTHKAQKENMPLLDKILSERVTLCDYELIVGDHGKRLLAFGKYAGRAGL 156

Qy        121 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
              :||||||||||||||||||||||| |:|| |||||||||| | ||||: ||| ||||:||
Db        157 VDFLHGLGQRYLSLGYSTPFLSLGASYMYSSLAAAKAAVISVGEEIASQGLPLGICPLVF 216

Qy        181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIF-QARNLSKQSQSTKRVFQLYGCVVTSR 239
              |||| |||| ||||||||||||||:  ||||:| : : :|:   |||||:|:|||::||:
Db        217 VFTGTGNVSLGAQEIFKLLPHTFVEPSKLPELFVKDKGISQNGISTKRVYQVYGCIITSQ 276

Qy        240 DIVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQ 299
              |:| ||||:: ||| ||||||||| |||||:|:|| ||:|||||||| || ||:  |||
Db        277 DMVEHKDPSKSFDKADYYAHPEHYNPVFHEKISPYTSVLVNCMYWEKSFPCLLSTKQLQD 336

Qy        300 LMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVD 359
              | : | ||||:|||||||||||||:|::| |: ||||::|| |||:|||:| ||:|:|||
Db        337 LTKKGLPLVGICDITCDIGGSIEFVNRATLIDSPFFRFNPSNNSYYDDMDGDGVLCMAVD 396

Qy        360 ILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMR 419
              ||||||:||||||||:|||  | ||||: : ::||::|:||||:: | || |||||||||
Db        397 ILPTEFAKEASQHFGDILSGFVGSLASMTEISDLPAHLKRACISYRGELTSLYEYIPRMR 456

Qy        420 NTMIDLAPAKTNP------------LPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGS 467
                       |:||               : :: |||||||||||||||||||:|| ||||
Db        457 ---------KSNPEEAQDNIIANGVSSQRTFNILVSLSGHLFDKFLINEALDMIEAAGGS 507

Qy        468 FHLVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLAN---EHGGDHDAGQEIELA 524
              ||| :||:||| |  |||||||||||   ||:|||||| |||   ::   |    :|  :
Db        508 FHLAKCELGQSADAESYSELEVGADDKRVLDQIIDSLTRLANPNEDYISPHREANKI--S 565

Qy        525 LKIGKVNEYETDVTIDKGGPK------ILILGAGRVCRPAAEFLASYPDICT-------Y 571
              |||||| : | ::   |  |:      :|||||||||||||:||||   | :       :
Db        566 LKIGKVQQ-ENEI---KEKPEMTKKSGVLILGAGRVCRPAADFLASVRTISSQQWYKTYF 621

Qy        572 GVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLP 631
              | |  :   :|||||||| |||:|||:|| :  | :|||:|  ||   ||||:||:||||
Db        622 GADSEEKTDVHVIVASLYLKDAKETVEGISDVEAVRLDVSDSESLLKYVSQVDVVLSLLP 681

Qy        632 ASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKM 691
              || || :|  |||||||:|||||||:  | | : || ||:||| ||||||||| :|: ||
Db        682 ASCHAVVAKTCIELKKHLVTASYVDDETSMLHENAKSAGITILGEMGLDPGIDPMMAKKM 741

Qy        692 IDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIH 751
              |::|   |||:|:|||||||||| |||||||||||||||||||:|:|:||| ||  |:  ||
Db        742 INDAPITKGKVKSFTSYCGGLPSRAAANNPLAYKFSWNPAGAIRAGQNPAKYKSNGDIIH 801

Qy        752 VDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEI 811
              ||| |||:|| | |:  |||||||  |||:||:||: |||  ||:||:| | |||||| |
Db        802 VDGKNLYDSAARFRVPNLPAFALECFPNRDSLVYGEHYGIESEATTIFRGTLRYEGFSMI 861

Qy        812 MVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIARL 871
              | |||| ||||:  : :|    | |:   |  :||     : | : |   |  ::::  |:
Db        862 MATLSKLGFFDSEANQVLSTGKRITFGALLSNILNK----DADNESEPLAG-EEEISKRI 916

Qy        872 LKLGCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVL 931
              :|||   :|| | |  ||| ||| :|| ::|  | | ||  |  ||:::||  |||||||
Db        917 IKLG--HSKETAAKAAKTIVFLGFNEEREVPSLCKSVFDATCYLMEEKLAYSGNEQDMVL 974

Qy        932 LHHEVEVEYPDGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTK 991
              ||||||||: : :  ||| ||||||| ::||::||||| |||||||||||||:::|::|:
Db        975 LHHEVEVEFLESKRIEKHTATLLEFGDIKNGQTTTAMAKTVGIPAAIGALLLIEDKIKTR 1034

Qy        992 GVIRPLQPEIYVPALEILESSGIKLVEKVE 1021
              ||:|||: |:|:|||:||:: ||||:|| |
Db       1035 GVLRPLEAEVYLPALDILQAYGIKLMEKAE 1064


RESULT   6
Q947M5
ID   Q947M5       PRELIMINARY;      PRT;   465 AA.
AC   Q947M5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MONOFUNCTIONAL LYSINE-KETOGLUTARATE REDUCTASE.
GN   LKR.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tang G., Zhu X., Galili G.;
RT   "Arabidopsis monofunctional lysine-ketoglutarate reductase.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295389; AAK97099.1; -.
SQ   SEQUENCE   465 AA;  51498 MW;  CE27C04A2F9C12ED CRC64;

  Query Match             31.6%;  Score 1670.5;  DB 10;  Length 465;
  Best Local Similarity   71.8%;  Pred. No. 1e-106;
  Matches  311;  Conservative   59;  Mismatches   52;  Indels   11;  Gaps    3;

Qy          1 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 60
              ||| || |||:   ::||:|||| :|||||| ||| |||||:|||:||||:|||||:|:|
Db         38 CAR-LLHGGKDRTGISRIVVQPSAKRIHHDALYEDVGCEISDDLSDCGLILGIKQPELEM 96

Qy         61 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 120
              || :|||||||||||||||||||||||| |||:| |||||||| ||| |||||:||||||
Db         97 ILPERAYAFFSHTHKAQKENMPLLDKILSERVTLCDYELIVGDHGKRLLAFGKYAGRAGL 156

Qy        121 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
              :||||||||||||||||||||||| |:|| |||||||||| | ||||: ||| ||||:||
Db        157 VDFLHGLGQRYLSLGYSTPFLSLGASYMYSSLAAAKAAVISVGEEIASQGLPLGICPLVF 216

Qy        181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIF-QARNLSKQSQSTKRVFQLYGCVVTSR 239
              |||| |||| ||||||||||||||:  ||||:| : : :|:   |||||:|:|||::||:
Db        217 VFTGTGNVSLGAQEIFKLLPHTFVEPSKLPELFVKDKGISQNGISTKRVYQVYGCIITSQ 276

Qy        240 DIVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQ 299
              |:| ||||:: ||| ||||||||| |||||:|:|| ||:|||||||| || ||:  |||
Db        277 DMVEHKDPSKSFDKADYYAHPEHYNPVFHEKISPYTSVLVNCMYWEKSFPCLLSTKQLQD 336

Qy        300 LMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVD 359
              | : | ||||:|||||||||||||:|::| |: ||||::|| |||:|||:| ||:|:|||
Db        337 LTKKGLPLVGICDITCDIGGSIEFVNRATLIDSPFFRFNPSNNSYYDDMDGDGVLCMAVD 396

Qy        360 ILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMR 419
              ||||||:||||||||:|||  | ||||: : ::||::|:||||:: | || |||||||||
Db        397 ILPTEFAKEASQHFGDILSGFVGSLASMTEISDLPAHLKRACISYRGELTSLYEYIPRMR 456

Qy        420 NTMIDLAPAKTNP 432
                       |:||
Db        457 ---------KSNP 460


RESULT   7
Q94BT4
ID   Q94BT4       PRELIMINARY;      PRT;   482 AA.
AC   Q94BT4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT4G33150/F4I10_80.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA   Bowser L., Carninci P., Dale J.M., Gibson H.A., Goldsmith A.D.,
RA   Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA   Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
```

```
RA   Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA   Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA   Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinozaki K.,
RA   Davis R.W., Theologis A., Ecker J.R.;
RT   "Arabidopsis cDNA clones.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY039906; AAK64010.1; -.
SQ   SEQUENCE   482 AA;  52655 MW;  5AB85A044BD7B942 CRC64;

  Query Match             28.0%;  Score 1481;  DB 10;  Length 482;
  Best Local Similarity   60.7%;  Pred. No. 1.2e-93;
  Matches  293;  Conservative   75;  Mismatches  101;  Indels   14;  Gaps    5;

Qy    546 ILILGAGRVCRPAAEFLASYPDIC------TY-GVDDHDADQIHVIVASLYQKDAEETVD 598
          :|||||||||||||:||||   |       || | |  :   :|||||||| |||:|||:
Db      7 VLILGAGRVCRPAADFLASVRTISSQQWYKTYVGADSEEKTDVHVIVASLYLKDAKETVE 66

Qy    599 GIENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDES 658
          || :  | :|||:|  ||   ||||:||:|||||| || :|  |||||||:|||||||:
Db     67 GISDVEAVRLDVSDSESLLKYVSQVDVVLSLLPASCHAVVAKTCIELKKHLVTASYVDDE 126

Qy    659 MSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAA 718
           | | : || ||:||| ||||||||||:|:||||::|| :|||:|:||||||||||||||
Db    127 TSMLHEKAKSAGITILGEMGLDPGIDHMMAMKMINDAHIKKGKVKSFTSYCGGLPSPAAA 186

Qy    719 NNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLP 778
          |||||||||||||||:|:|:||| ||  |: ||||| |||:|| | |:  |||||||  |
Db    187 NNPLAYKFSWNPAGAIRAGQNPAKYKSNGDIIHVDGKNLYDSAARFRVPNLPAFALECFP 246

Qy    779 NRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYK 838
          ||:||:||: |||  ||:||:| | |||||| || |||| ||||:  : :|    | |:
Db    247 NRDSLVYGEHYGIESEATTIFRGTLRYEGFSMIMATLSKLGFFDSEANQVLSTGKRITFG 306

Qy    839 GFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHEE 898
            |  :||     : | : |   | ::::  |::|||   :|| | |  ||| ||| :||
Db    307 ALLSNILNK----DADNESEPLAG-EEEISKRIIKLG--HSKETAAKAAKTIVFLGFNEE 359

Qy    899 TQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEFGK 958
           ::|  | | ||  |  ||:::||  |||||||||||||||: : :  ||| |||||||
Db    360 REVPSLCKSVFDATCYLMEEKLAYSGNEQDMVLLHHEVEVEFLESKRIEKHTATLLEFGD 419

Qy    959 VENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSGIKLVE 1018
          ::||::||||| |||||||||||||:::|::|:||:|||: |:|:|||:||:: ||||:|
Db    420 IKNGQTTTAMAKTVGIPAAIGALLLIEDKIKTRGVLRPLEAEVYLPALDILQAYGIKLME 479

Qy   1019 KVE 1021
          | |
Db    480 KAE 482


RESULT    8
Q9VLX0
ID   Q9VLX0      PRELIMINARY;      PRT;   972 AA.
AC   Q9VLX0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG7144 PROTEIN.
GN   BEST:CK02318 OR CG7144.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
```

```
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003618; AAF52559.1; -.
DR   FlyBase; FBgn0025687; BEST:CK02318.
DR   FlyBase; FBgn0030715; Or13a.
SQ   SEQUENCE   972 AA;  108214 MW;  7DA8002E16DBC1E9 CRC64;

  Query Match             27.7%;  Score 1466.5;  DB 5;  Length 972;
  Best Local Similarity   33.4%;  Pred. No. 3.8e-92;
  Matches  343;  Conservative  183;  Mismatches  318;  Indels  183;  Gaps   22;

Qy     17 RIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQMILSDRAYAFFSHTHKA 76
          ::||||| || :    |  ||  | ||:|:  :| |:||  :  ::  : | ||||| ||
Db     98 KVIVQPSNRRAYPMQAYMQAGAHIQEDISDASVIFGVKQVPIDALIPGKTYCFFSHTIKA 157

Qy     77 QKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRYLSLGY 136
          |: |||||| |||::: | ||| |: : | | :||||:|| ||::: |||:| | |:||:
Db    158 QESNMPLLDAILEKKIRLIDYERIIDERGARQVAFGKYAGVAGMVNILHGIGLRLLALGH 217

Qy    137 STPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVSQGAQEIF 196
           |||: :| :| | : : |: |:     ||:   :|  | |: ||||| |||||||||:|
Db    218 HTPFMHIGPAHNYRNSSMARQAIRDCGYEISLGMMPKSIGPLTFVFTGSGNVSQGAQEVF 277

Qy    197 KLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDIVSHKDPTRQFDKGDY 256
            ||  :|  | | :: :  | :|          |||| |:  | :  ::    ||  :|
Db    278 SELPIEYVPPEMLRKVAEHGNQNK----------LYGCEVSRSDHLERRE-GGGFDAKEY 326

Qy    257 YAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLM----------ETGCP 306
             || |   |  :|||||||||| :||    | |:::   : |:            | |
Db    327 DEFPERYISTFSTKIAPYASVIVNGIYWAVGSPKLISIPDAKNLLRPANTPWLPTSKGSP 386

Qy    307 -----LVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDIL 361
               :: :|||: | ||||||:|: |:|: ||  ||  :|      :| ||:  ::| :
Db    387 ALPHRMLAICDISADPGGSIEFMNECTTIDTPFCLYDADRNKDTKSFKGPGVLVCSIDNM 446

Qy    362 PTEFSKEASQHFGNILSRLVASLASVKQPAELP------SY-LRRACIAHAGRLTPLYEY 414
          ||:  :|::  || :|:  |  :  :|  |: |      || :: | ||  |:||  ::|
Db    447 PTQLPRESTDLFGELLAPHVHDI--IKSDAKKPLAEENFSYPIQSAIIASNGQLTEGFQY 504

Qy    415 IPRMRNTMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCE 474
```

```
           |  :|
Db     505 IQELR------------------------------------------------------- 509

Qy     475 VGQSTDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYE 534
             :|    |  ::| |: ::
Db     510 --ESQSHRSRHKME-GSSES---------------------------------------- 526

Qy     535 TDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAE 594
                ||   |:|:|||| |  |  |:|             |    : : | |  :::|:
Db     527 -----DK---KVLVLGAGMVSAPLVEWL-------------HREKDVSITVCSQVKEEAD 565

Qy     595 ETVDGIENTTATQLDVAD-IGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTAS 653
                     :  ||| :  | | :|  : :||:|||| | |  :|  |:    ||||||
Db     566 RLAQQYAGVDSVYLDVNESTGHLQELCGRADVVVSLLPYSLHGMVARYCVAEGTHMVTAS 625

Qy     654 YVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLP 713
           |::: :| | : ||  ||||: |:|||||||||:::: | |   :   :::| |||||||
Db     626 YLNDEISGLHEEAKAKGVTIMNEVGLDPGIDHLLALECIHEVQDKGAVVESFVSYCGGLP 685

Qy     714 SPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFA 773
           :|  :|| | |||||:| | | :  : | |   |: : : |     |: |  |  || ||
Db     686 APEHSNNALRYKFSWSPRGVLLNTLSAAKYLSQGQIVEISGGGELMSSPR-SLDFLPGFA 744

Qy     774 LEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTS 833
           ||  |||:|  ||:|||: ::  |: | | ||:|||| :  :   |  |   | ||  :
Db     745 LEGFPNRDSTKYGNLYGLGRDVHTLLRGTIRYKGFSESIKPMQLLGLIDPEPHALLHPSG 804

Qy     834 RP-TYKGFLDELLN-NISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIK 891
              |::  :  |:  : |||           : ::|  :| :          :  |  |:
Db     805 PDVTWRQLVIHLMGMSDSTI-----------FYENLKQKLTE---------RIGDVDGIE 844

Qy     892 FLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQA 951
            ||| ::| : |  ::| | :   : :|:|:  :|:|:|:| ||| : :|||:  |:
Db     845 SLGLLDDTPVVK-LNTPLDTLSHYLSKRLAFERDERDLVVLRHEVGIRWPDGR-REERGI 902

Qy     952 TLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILES 1011
             : :|: : | |  |||:||| |||| | ::|  ::| :||: |  |:|| | |: | |
Db     903 NFVVYGQPQ-GHS--AMAMTVGKPAAIAAKMILDGEIQERGVLLPFTPDIYRPMLQRLRS 959

Qy    1012 SGIKLVE 1018
            |:   |
Db     960 EGLTATE 966

RESULT   9
O44503
ID   O44503     PRELIMINARY;     PRT;   934 AA.
AC   O44503;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   R02D3.1 PROTEIN.
GN   R02D3.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Wu X., Antoniou B.;
RT   "The sequence of C. elegans cosmid R02D3.";
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF038615; AAB94142.1; -.
SQ   SEQUENCE   934 AA;  103751 MW;  CF985E127DDA69D5 CRC64;

  Query Match             27.0%;  Score 1429.5;  DB 5;  Length 934;
  Best Local Similarity   32.5%;  Pred. No. 1.2e-89;
  Matches  334;  Conservative  190;  Mismatches  321;  Indels  183;  Gaps   24;

Qy      10 KNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQMILSDRAYAF 69
           |:|  || :::||| || :    |  ||  : |||||  :|: :|   :  :: :: |||
Db      58 KSG--VN-VLIQPSNRRAYPIQDYISAGAIVREDLSEAHIIMSVKSVPIDQLIPNKTYAF 114

Qy      70 FSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQ 129
           |||| |||::|| :|| ||:  : | ||| |  | ||| : |||:|| || || |||:|
Db     115 FSHTIKAQQDNMEMLDTILQRNIRLLDYEKICDDKGKRLVMFGKWAGNAGFIDILHGIGL 174

Qy     130 RYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVS 189
           | |:||::|||| :| :| |     |  |:  :  |||   :|  : |::||||| ||||
Db     175 RLLALGHNTPFLHMGLAHNYNDSHMAINALRDIGYEIALDRMPRSLGPLIFVFTGSGNVS 234

Qy     190 QGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRD--IVSHKDP 247
           |||:|:|: ||| :||   ||:: |   |:|          :|||||| :|  :  |  |
Db     235 QGARELFEHLPHEYVDVATLPKVAQKGQLNK----------VYGCVVTRKDHFVPKHGGP 284

Qy     248 TRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLM------ 301
              ||| ::   |: ||  |   ||||||||:| :||: : | |: :   : |:
Db     285 ---FDKKEFEQFPDRYTSKFATEIAPYASVIINGVYWDAQSPRLITIPDAKNLLTPVQRY 341

Qy     302 -ETGCP-----LVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGA-GVV 354
              |||     |: :|||: | |||:||: : |:|::||  ||   |:  |  : | | :
Db     342 DTPGCPTLPHHLIALCDISADPGGSVEFMRECTTIDKPFAIYDADFNTSSDSFDAASGCL 401

Qy     355 CLAVDILPTEFSKEASQHFGNILSRLVASLASVKQP-----AELPSYLRRACIAHAGRLT 409
             ::| :| :   ||:: |||:|   :  | :          :  : :: | |   |:||
Db     402 VCSIDNMPAQMPIEATEQFGNLLYPWLWDLLNTSNDQHFDRLQCRTEIKNAIITDQGKLT 461

Qy     410 PLYEYIPRMRNTMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFH 469
           | :||| ::|      | | :  :
Db     462 PNFEYIAQLRKDKAASASANSRVM------------------------------------ 485

Qy     470 LVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGK 529
                                                     ||
Db     486 ------------------------------------------GG---------------- 487

Qy     530 VNEYETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLY 589
                   | ||   ::|:|||| |  | |:| :   |             ::: ||:
Db     488 --------TTDK---RVLLLGAGMVSGPFADFYSKQKD-------------VNLTVATES 523

Qy     590 QKDAEETVDGIENTTATQLDVA-DIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKH 648
           |:| :       |  :  :|:| :  ::  |: : ::|:|||| :||  :| :||  ::
Db     524 QRDGQRLCTS-PNIQSVVVDIARESHTMERLIREHDLVVSLLPFNFHPLVAKMCISNQRD 582

Qy     649 MVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSY 708
           |||:|||   :  | :||||| |||: | ||||||||:::|:  |:     |:| :| |:
Db     583 MVTSSYVSPELEALDKAAKDADVTIMNEAGLDPGIDHMLAMECFDDIKEHGGRITSFESF 642

Qy     709 CGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRE 768
           |||||:|  ::||| |||||:| | | :  ||| |   |: : |   :: ::   : :
Db     643 CGGLPAPEWSDNPLRYKFSWSPKGVLTALMNPAKYLKNGKIVEVPSGSVVDNL--IDIDF 700

Qy     769 LPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPL 828
```

```
           :|   |   |||:|  | |:||:  :  || | | ||:|| : :  |   |     |
Db     701 MPGLNLIGFPNRDSTKYSDVYGLGNDCKTIIRGTLRYQGFVDTVKALHSVGLLSGDNIDS 760

Qy     829 LQDTSRP--TYKGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKT 886
                 |  |:|    ||: ::|  |  |||     : | |   :        :::  |
Db     761 FTSNIGPDLTWK----ELIASLS--NQKLDI-----FPDSLRHII-------EEKVGGKG 802

Qy     887 VKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPA 946
           :  :: |||  :  : :   :| | : | : : :|:  :| |:|:|:|::  : | |  :
Db     803 LSALENLGLFSDKVVDRH-GTPIDTLAQYLAKILAFKDHESDLVVLNHDIGAQLPGGN-S 860

Qy     947 EKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPAL 1006
           |:|: :|:::|   |    :||| |||   ||  : ::| |::|  |:  ||:  |:| |||
Db     861 ERHRISLVQYG---NPNGFSAMARTVGYTTAIVSHMVLNNEIQRAGIQRPILKEVYRPAL 917

Qy    1007 EILESSGI 1014
           : |   ||
Db     918 KRLRDFGI 925


RESULT  10
Q9Z1I9
ID   Q9Z1I9      PRELIMINARY;      PRT;   926 AA.
AC   Q9Z1I9;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LYSINE KETOGLUTARATE REDUCTASE/SACCHAROPINE DEHYDROGENASE.
GN   LORSDH OR LKR/SDH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6CBAF1/J; TISSUE=LIVER;
RX   MEDLINE=20036337; PubMed=10567240;
RA   Papes F., Kemper E.L., Cord-Neto G., Langone F., Arruda P.;
RT   "Lysine degradation through the saccharopine pathway in mammals:
RT   involvement of both bifunctional and monofunctional lysine-degrading
RT   enzymes in mouse.";
RL   Biochem. J. 344:555-563(1999).
DR   EMBL; AJ224761; CAA12114.1; -.
DR   MGD; MGI:1353573; Lorsdh.
DR   InterPro; IPR002016; Peroxidase.
DR   PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
SQ   SEQUENCE   926 AA;  102946 MW;  E01118983B27A2B5 CRC64;

  Query Match             26.3%;  Score 1388.5;  DB 11;  Length 926;
  Best Local Similarity   31.7%;  Pred. No. 8.2e-87;
  Matches  327;  Conservative  186;  Mismatches  316;  Indels  201;  Gaps   22;

Qy      17 RIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQMILSDRAYAFFSHTHKA 76
           ::::||| ||  || :|  ||  : ||::|  ||:|:|:|  : ::| : ||||||| ||
Db      56 KVLIQPSNRRAIHDKEYVRAGGILQEDITEACLILGVKRPPEEKLMSKKTYAFFSHTIKA 115

Qy      77 QKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRYLSLGY 136
           |: || |||::|:: : | ||| :|   | | :|||::|| ||:|: |||:| | |:||:
Db     116 QEANMNLLDEVLKQEIRLIDYEKMVDHRGSRIVAFGQWAGVAGMINILHGMGLRLLALGH 175

Qy     137 STPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVSQGAQEIF 196
            |||: || :| | : : |  |:     ||:   :|  | |: ||||| ||||:||||:|
Db     176 HTPFMHLGMAHNYRNSSQAVQAIRDAGYEISLGLMPKSIGPLTFVFTGTGNVSKGAQEVF 235

Qy     197 KLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDIVSHKDPTRQ----FD 252
             ||  :|:  :| |: :  :| |          :|| |::      |    |:    :|
Db     236 NELPCEYVEPHELREVSKTGDLRK----------VYGTVLS-----RHHHLVRKTDGVYD 280

Qy     253 KGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLM---------ET 303
             :|  :|| ||  |:  |||| : ::| :|||:  | ||     | |:
Db     281 PVEYEKYPERYTSRFNTDIAPYTTCLINGIYWEQNTPRLLTRQDAQSLLVPVKSSVVPVE 340
```

```
Qy     304 GCP-----LVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAV 358
           |||     || :|||: | ||||:|: : |:|||||  ||  :   || :||:|::  ::
Db     341 GCPELPHKLVAICDISADTGGSIDFMTECTTIERPFCMYDADQQIIHDSVEGSGILMCSI 400

Qy     359 DILPTEFSKEASQHFGNILSRLVAS--LASVKQPAELPSY---LRRACIAHAGRLTPLYE 413
           | || :   ||:::||::|   |    |:   || |  ::   :| | |   | ||  |:
Db     401 DNLPAQLPIEATEYFGDMLYPYVEEMLLSDASQPLESQNFSPVVRDAVITSNGLLTDKYK 460

Qy     414 YIPRMRNTMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRC 473
           || ::| :
Db     461 YIQKLRES---------------------------------------------------- 468

Qy     474 EVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEY 533
                                                     :  |  : :: |
Db     469 -----------------------------------------RERIQFLSMSTK------- 480

Qy     534 ETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDI-CTYGVDDHDADQIHVIVASLYQKD 592
                      |:|:||:| |  |  |:|:   :|  | | |         :   : |
Db     481 ----------KKVLVLGSGYVSGPVLEYLSRDNNIEITLGSD---------MTNQMQQLS 521

Qy     593 AEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTA 652
            :  :: :  |   |        |  ||   ::||||||   |  :|  ||| : :||||
Db     522 KKYNINPVSLTVGKQ-----EAKLQSLVESQDLVISLLPYVLHPVVAKACIESRVNMVTA 576

Qy     653 SYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGL 712
           ||:  :|  | ::  |||:|:: |:|||||:||:::|: || |      :::: ||||||
Db     577 SYITPAMKELEKSVDDAGITVIGELGLDPGLDHMLAMETIDTAKELGATVESYVSYCGGL 636

Qy     713 PSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAF 772
           |:|  ::||| |||||:| | | :   || |   |: ::| |   : ::    :   |
Db     637 PAPEHSDNPLRYKFSWSPVGVLMNIMQPASYLLNGKVVNVTGGVSFLNSV-TPMDYFPGL 695

Qy     773 ALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDT 832
            ||  |||:|: | ::||||  | |: | | ||:|:|: :    | |  :   :| |:
Db     696 NLEGYPNRDSIKYAEIYGIS-SAHTLLRGTLRYKGYSKALNGFVKLGLINREAYPALRPE 754

Qy     833 SRP-TYKGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAV------- 884
           : | |:|  | :|:                      |:|     | | ||:
Db     755 ANPLTWKQLLCDLVG---------------------ISR--SSPCEKLKEVVFTKLGGDN 791

Qy     885 KTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQ 944
              :: ::|||  : |:|:   |  |   : :  :::||  |:||:::     : :| |
Db     792 TQLEAAEWLGLLGDEQVPQ-AESIVDAFSKHLVSKLSYGPEEKDMIVMRDSFGIRHPSGH 850

Qy     945 PAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVP 1004
            ||     |: :|   || |  ||| |||:| |: | :||  ::: ||:: |   ||| |
Db     851 -AENKTIDLVVYGDF-NGFS--AMAKTVGLPTAMAAKMLLDGEIEAKGLMGPFTKEIYGP 906

Qy    1005 ALEILESSGI 1014
            || ::: ||
Db     907 ILERIKAEGI 916


RESULT  11
Q99K67
ID   Q99K67      PRELIMINARY;      PRT;   926 AA.
AC   Q99K67;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LYSINE OXOGLUTARATE REDUCTASE, SACCHAROPINE DEHYDROGENASE.
GN   LORSDH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC005420; AAH05420.1; -.
DR   MGD; MGI:1353573; Lorsdh.
DR   InterPro; IPR002016; Peroxidase.
```

```
DR   PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
SQ   SEQUENCE   926 AA;  102974 MW;  5B4369C51F7D1D53 CRC64;

  Query Match             26.2%;  Score 1384.5;  DB 11;  Length 926;
  Best Local Similarity   31.7%;  Pred. No. 1.5e-86;
  Matches  327;  Conservative  185;  Mismatches  317;  Indels  201;  Gaps   22;

Qy         17 RIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQMILSDRAYAFFSHTHKA 76
              ::::||| ||  || :|  ||  : ||::|  ||:|:|:|  : ::| : ||||||| ||
Db         56 KVLIQPSNRRAIHDKEYVRAGGILQEDITEACLILGVKRPPEEKLMSKKTYAFFSHTIKA 115

Qy         77 QKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRYLSLGY 136
              |: || |||::|:: : | ||| :|   | | :|||::|| ||:|: |||:| | |:||:
Db        116 QEANMNLLDEVLKQEIRLIDYEKMVDHRGSRIVAFGQWAGVAGMINILHGMGLRLLALGH 175

Qy        137 STPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVSQGAQEIF 196
               |||: || :| | : : |  ||     ||:   :|  | |: ||||| ||||:||||:|
Db        176 HTPFMHLGMAHNYRNSSQAVQAVRDAGYEISLGLMPKSIGPLTFVFTGTGNVSKGAQEVF 235

Qy        197 KLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDIVSHKDPTRQ----FD 252
                ||  :|:  :| |: :  :| |          :|| |::      |    |:    :|
Db        236 NELPCEYVEPHELREVSKTGDLRK----------VYGTVLS-----RHHHLVRKTDGVYD 280

Qy        253 KGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLM---------ET 303
                :|  :|| ||  |:  |||| : ::| :|||:  | ||     | |:
Db        281 PVEYEKYPERYTSRFNTDIAPYTTCLINGIYWEQNTPRLLTRQDAQSLLVPVKSSVVPVE 340

Qy        304 GCP-----LVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAV 358
              |||     || :|||: | ||||:|: : |:|||||  ||  :   || :||:|::  ::
Db        341 GCPELPHKLVAICDISADTGGSIDFMTECTTIERPFCMYDADQQIIHDSVEGSGILMCSI 400

Qy        359 DILPTEFSKEASQHFGNILSRLVAS--LASVKQPAELPSY---LRRACIAHAGRLTPLYE 413
              | || :   ||:::||::|   |    |:   || |  ::   :| | |   | ||  |:
Db        401 DNLPAQLPIEATEYFGDMLYPYVEEMLLSDASQPLESQNFSPVVRDAVITSNGLLTDKYK 460

Qy        414 YIPRMRNTMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRC 473
              || ::| :
Db        461 YIQKLRES---------------------------------------------------- 468

Qy        474 EVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEY 533
                                                        :  | : :: |
Db        469 -----------------------------------------RERIQFLSMSTK------- 480

Qy        534 ETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDI-CTYGVDDHDADQIHVIVASLYQKD 592
                         |:|:||:| |  |  |:|:   :|  | | |         :   : |
Db        481 ----------KKVLVLGSGYVSGPVLEYLSRDNNIEITLGSD---------MTNQMQQLS 521

Qy        593 AEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTA 652
               :  :: :  |   |        |  ||   ::||||||   |  :|  ||| : :||||
Db        522 KKYNINPVSLTVGKQ-----EAKLQSLVESQDLVISLLPYVLHPVVAKACIESRVNMVTA 576

Qy        653 SYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGL 712
              ||:  :|  | ::  |||:|:: |:|||||:||:::|: || |      :::: ||||||
Db        577 SYITPAMKELEKSVDDAGITVIGELGLDPGLDHMLAMETIDTAKELGATVESYVSYCGGL 636

Qy        713 PSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAF 772
              |:|  ::||| |||||:| | | :   || |   |: ::| |   : ::    :   |
Db        637 PAPEHSDNPLRYKFSWSPVGVLMNIMQPASYLLNGKVVNVTGGVSFLNSV-TPMDYFPGL 695

Qy        773 ALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDT 832
               ||  |||:|: | ::||||  | |: | | ||:|:|: :    | |  :   :| |:
Db        696 NLEGYPNRDSIKYAEIYGIS-SAHTLLRGTLRYKGYSKALNGFVKLGLINREAYPALRPE 754

Qy        833 SRP-TYKGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAV------- 884
              : | |:|  | :|:                      |:|     | | ||:
Db        755 ANPLTWKQLLCDLVG---------------------ISR--SSPCEKLKEVVFTKLGGDN 791

Qy        885 KTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQ 944
                ::  ::|||  : |:|:   |  |   : :  :::||  |:||:::     : :| |
Db        792 TQLEAAEWLGLLGDEQVPQ-AESIVDAFSKHLVSKLSYGPEEKDMIVMRDSFGIRHPSGH 850

Qy        945 PAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVP 1004
                |     |: :|   || |  ||| |||:| |: | :||  ::: ||:: |   ||| |
Db        851 -LENKTIDLVVYGDF-NGFS--AMAKTVGLPTAMAAKMLLDGEIEAKGLMGPFTKEIYGP 906

Qy       1005 ALEILESSGI 1014
               || ::: ||
Db        907 ILERIKAEGI 916


RESULT  12
Q9UDR5
ID   Q9UDR5      PRELIMINARY;      PRT;   926 AA.
AC   Q9UDR5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LYSINE KETOGLUTARATE REDUCTASE/SACCHAROPINE DEHYDROGENASE (ALPHA-
DE   AMINOADIPATE SEMIALDEHYDE SYNTHASE).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99063792; PubMed=9847074;
RA   Sulston J.E., Waterston R.;
RT   "Toward a complete human genome sequence.";
RL   Genome Res. 8:1097-1108(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Drone K., Le T.;
RT   "The sequence of Homo sapiens PAC clone RP5-1049N15.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Waterston R.H.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Waterston R.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Sacksteder K.A., Biery B.J., Morrell J.C., Goodman B.K.,
RA   Geisbrecht B.V., Cox R.P., Gould S.J., Geraghty M.T.;
RT   "Identification of the Alpha-Aminoadipic Semialdehyde Synthase Gene
RT   which is Defective in Familial Hyperlysinemia.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006020; AAF03526.1; -.
DR   EMBL; AF229180; AAF44328.1; -.
DR   InterPro; IPR002016; Peroxidase.
DR   PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
SQ   SEQUENCE   926 AA;  102131 MW;  CB4194014351A18D CRC64;

  Query Match             25.8%;  Score 1363.5;  DB 4;  Length 926;
  Best Local Similarity   32.2%;  Pred. No. 4.3e-85;
  Matches  328;  Conservative  175;  Mismatches  336;  Indels  181;  Gaps   20;

Qy         17 RIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQMILSDRAYAFFSHTHKA 76
              ::::||| ||  ||  |  ||  : ||:||  ||:|:|:|  : ::| : ||||||| ||
Db         56 KVLIQPSNRRAIHDKDYVKAGGILQEDISEACLILGVKRPPEEKLMSRKTYAFFSHTIKA 115

Qy         77 QKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRYLSLGY 136
              |: || |||:||:: : | ||| :|   | | :|||::|| ||:|: |||:| | |:||:
Db        116 QEANMGLLDEILKQEIRLIDYEKMVDHRGVRVVAFGQWAGVAGMINILHGMGLRLLALGH 175

Qy        137 STPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVSQGAQEIF 196
               |||: :| :| | : : |  ||     ||:   :|  | |: ||||| ||||:||| ||
Db        176 HTPFMHIGMAHNYRNSSQAVQAVRDAGYEISLGLMPKSIGPLTFVFTGTGNVSKGAQAIF 235

Qy        197 KLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDIVSHKDPTRQFDKGDY 256
                ||  :|:  :| |: |  :| |          :|| |::    :  |     :|  :|
Db        236 NELPCEYVEPHELKEVSQTGDLRK----------VYGTVLSRHHHLVRKTDA-VYDPAEY 284
```

```
Qy        257 YAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLME---------TGCP- 306
                ||| |   |:  |||| : ::| :|||:  | ||     | |:           |||
Db        285 DKHPERYISRFNTDIAPYTTCLINGIYWEQNTPRLLTRQDAQSLLAPGKFSPAGVEGCPA 344

Qy        307 ----LVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDILP 362
                  || :|||: | ||||||: : |:|| ||  ||  ::  || :||:|::  ::| ||
Db        345 LPHKLVAICDISADTGGSIEFMTECTTIEHPFCMYDADQHIIHDSVEGSGILMCSIDNLP 404

Qy        363 TEFSKEASQHFGNILSRLVAS--LASVKQPAELPSY---LRRACIAHAGRLTPLYEYIPR 417
               :   ||:: ||::|   |    |:   || |  ::   :| | |   | |   |:||
Db        405 AQLPIEATECFGDMLYPYVEEMILSDATQPLESQNFSPVVRDAVITSNGTLPDKYKYIQT 464

Qy        418 MRNTMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQ 477
              :| :
Db        465 LRES-------------------------------------------------------- 468

Qy        478 STDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYETDV 537
                                                    :  | : :  :
Db        469 -------------------------------------RERAQSLSMGTR----------- 480

Qy        538 TIDKGGPKILILGAGRVCRPAAEFLASYPDI-CTYGVDDHDADQIHVIVASLYQKDAEET 596
                     |:|:||:| :  |  |:|:   :|  | | |                |:  |
Db        481 ------RKVLVLGSGYISEPVLEYLSRDGNIEITVGSD----------------MKNQIEQ 519

Qy        597 VDGIENTTATQLDVA-DIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYV 655
              :    |     :|:      |  ||:: ::||||||   |  :|  ||  | :||||||:
Db        520 LGKKYNINPVSMDICKQEEKLGFLVAKQDLVISLLPYVLHPLVAKACITNKVNMVTASYI 579

Qy        656 DESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSP 715
                ::  | :: :|||:||: |:|||||:||:::|: ||:|      |::: |||||||:|
Db        580 TPALKELEKSVEDAGITIIGELGLDPGLDHLMAMETIDKAKEVGATIESYISYCGGLPAP 639

Qy        716 AAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALE 775
                :|||| |||||:| | | :    | |   |: ::| |   :  |    :   |   ||
Db        640 EHSNNPLRYKFSWSPVGVLMNVMQSATYLLDGKVVNVAGGISFLDAV-TSMDFFPGLNLE 698

Qy        776 HLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRP 835
                |||:|  | ::||||  | |: | | ||:|: : :    | |  :    |  :  : |
Db        699 GYPNRDSTKYAEIYGIS-SAHTLLRGTLRYKGYMKALNGFVKLGLINREALPAFRPEANP 757

Qy        836 -TYKGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLG 894
               |:|  | :|:           |  |  :|    | | |||    :      ::  ::||
Db        758 LTWKQLLCDLVG----------ISPSSEHDVLKEAVLKKLGGDNTQ------LEAAEWLG 801

Qy        895 LHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLL 954
              |  : |:|:   |  | : : :  :::||  |:||:::     : :| |   |     |:
Db        802 LLGDEQVPQ-AESILDALSKHLVMKLSYGPEEKDMIVMRDSFGIRHPSGH-LEHKTIDLV 859

Qy        955 EFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSGI 1014
               :| : || |  ||| |||:| |: | :||  ::  ||:: |   ||| | || ::: ||
Db        860 AYGDI-NGFS--AMAKTVGLPTAMAAKMLLDGEIGAKGLMGPFSKEIYGPILERIKAEGI 916


RESULT  13
O95462
ID   O95462     PRELIMINARY;      PRT;   926 AA.
AC   O95462;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LYSINE-KETOGLUTARATE REDUCTASE /SACCHAROPINE DEHYDROGENASE.
GN   LKR/SDH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Papes F., Kemper E.L., Cord-Neto G., Langone F., Arruda P.;
RT   "Cloning and expression analysis of the LKR/SDH gene in human
RT   tissues.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ007714; CAA07619.2; -.
DR   InterPro; IPR002016; Peroxidase.
DR   PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
SQ   SEQUENCE   926 AA;  102147 MW;  DBFD9E17EF5CB680 CRC64;

  Query Match             25.8%;  Score 1362.5;  DB 4;  Length 926;
  Best Local Similarity   32.2%;  Pred. No. 5.1e-85;
  Matches  328;  Conservative  174;  Mismatches  337;  Indels  181;  Gaps   20;

Qy         17 RIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQMILSDRAYAFFSHTHKA 76
              ::::||| ||  ||  |  ||  : ||:||  ||:|:|:|  : ::| : ||||||| ||
Db         56 KVLIQPSNRRAIHDKDYVKAGGILQEDISEACLILGVKRPPEEKLMSRKTYAFFSHTIKA 115

Qy         77 QKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRYLSLGY 136
              |: || |||:||:: : | ||| :|   | | :|||::|| ||:|: |||:| | |:||:
Db        116 QEANMGLLDEILKQEIRLIDYEKMVDHRGVRVVAFGQWAGVAGMINILHGMGLRLLALGH 175

Qy        137 STPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVSQGAQEIF 196
               |||: :| :| | : : |  ||     ||:   :|  | |: ||||| ||||:||| ||
Db        176 HTPFMHIGMAHNYRNSSQAVQAVRDAGYEISLGLMPKSIGPLTFVFTGTGNVSKGAQAIF 235

Qy        197 KLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDIVSHKDPTRQFDKGDY 256
                ||  :|:  :| |: |  :| |          :|| |::    :  |     :|  :|
Db        236 NELPCEYVEPHELKEVSQTGDLRK----------VYGTVLSRHHHLVRKTDA-VYDPAEY 284

Qy        257 YAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLME---------TGCP- 306
                ||| |   |:  |||| : ::| :|||:  | ||     | |:           |||
Db        285 DKHPERYISRFNTDIAPYTTCLINGIYWEQNTPRLLTRQDAQSLLAPGKFSPAGVEGCPA 344

Qy        307 ----LVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDILP 362
                  || :|||: | ||||||: : |:|| ||  ||  ::  || :||:|::  ::| ||
Db        345 LPHKLVAICDISADTGGSIEFMTECTTIEHPFCMYDADQHIIHDSVEGSGILMCSIDNLP 404

Qy        363 TEFSKEASQHFGNILSRLVAS--LASVKQPAELPSY---LRRACIAHAGRLTPLYEYIPR 417
               :   ||:: ||::|   |    |:   || |  ::   :| | |   | |   |:||
Db        405 AQLPIEATECFGDMLYPYVEEMILSDATQPLESQNFSPVVRDAVITSNGTLPDKYKYIQT 464

Qy        418 MRNTMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQ 477
              :| :
Db        465 LRES-------------------------------------------------------- 468

Qy        478 STDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYETDV 537
                                                    :  | : :  :
Db        469 -------------------------------------RERAQSLSMGTR----------- 480

Qy        538 TIDKGGPKILILGAGRVCRPAAEFLASYPDI-CTYGVDDHDADQIHVIVASLYQKDAEET 596
                     |:|:||:| :  |  |:|:   :|  | | |                |:  |
Db        481 ------RKVLVLGSGYISEPVLEYLSRDGNIEITVGSD----------------MKNQIEQ 519

Qy        597 VDGIENTTATQLDVA-DIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYV 655
              :    |     :|:      |  ||:: ::||||||   |  :|  ||  | :||||||:
Db        520 LGKKYNINPVSMDICKQEEKLGFLVAKQDLVISLLPYVLHPLVAKACITNKVNMVTASYI 579

Qy        656 DESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSP 715
                ::  | :: :|||:||: |:|||||:||:::|: ||:|      |::: |||||||:|
Db        580 TPALKELEKCVEDAGITIIGELGLDPGLDHLMAMETIDKAKEVGATIESYISYCGGLPAP 639

Qy        716 AAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALE 775
                :|||| |||||:| | | :    | |   |: ::| |   :  |    :   |   ||
Db        640 EHSNNPLRYKFSWSPVGVLMNVMQSATYLLDGKVVNVAGGISFLDAV-TSMDFFPGLNLE 698

Qy        776 HLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRP 835
                |||:|  | ::||||  | |: | | ||:|: : :    | |  :    |  :  : |
Db        699 GYPNRDSTKYAEIYGIS-SAHTLLRGTLRYKGYMKALNGFVKLGLINREALPAFRPEANP 757

Qy        836 -TYKGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLG 894
               |:|  | :|:           |  |  :|    | | |||    :      ::  ::||
Db        758 LTWKQLLCDLVG----------ISPSSEHDVLKEAVLKKLGGDNTQ------LEAAEWLG 801

Qy        895 LHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLL 954
              |  : |:|:   |  | : : :  :::||  |:||:::     : :| |   |     |:
```

```
Db     802 LLGDEQVPQ-AESILDALSKHLVMKLSYGPEEKDMIVMRDSFGIRHPSGH-LEHKTIDLV 859

Qy     955 EFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSGI 1014
            :| : || |  ||| |||:| |: | :||  ::  ||:: |   ||| | || ::: ||
Db     860 AYGDI-NGFS--AMAKTVGLPTAMAAKMLLDGEIGAKGLMGPFSKEIYGPILERIKAEGI 916

RESULT  14
Q9C2G6
ID   Q9C2G6       PRELIMINARY;       PRT;    448 AA.
AC   Q9C2G6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PROBABLE SACCHAROPINE REDUCTASE (LYS3).
GN   93G11.270.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL513443; CAC28679.1; -.
SQ   SEQUENCE    448 AA;  49059 MW;  1EF38083F6CC1215 CRC64;

  Query Match             14.1%;  Score 746.5;  DB 3;  Length 448;
  Best Local Similarity   35.8%;  Pred. No. 4.3e-43;
  Matches  173;  Conservative   94;  Mismatches  165;  Indels   51;  Gaps   11;

Qy     545 KILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGIENTT 604
           |:|:|||| | ||  : |:                 | | ||      |::  :|::| |
Db       6 KVLMLGAGFVTRPTLDVLSE--------------AGIPVTVACRTLASAQKLSEGVKNAT 51

Qy     605 ATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQ 664
              ||| :  :|   |:: ::||||:| :||| :    |  |||:|| |||  :|  |
Db      52 PISLDVTNDEALDAEVAKHDLVISLIPYTFHATVIKSAIRQKKHVVTTSYVSPAMMELDA 111

Qy     665 AAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAY 724
            || ||:|:: |:||||||||| ::| ||| |   ||| :| |||||||:|  ::||| |
Db     112 EAKAAGITVMNEIGLDPGIDHLYAIKTIDEVHQAGGKILSFLSYCGGLPAPEDSDNPLGY 171

Qy     725 KFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSLI 784
           ||||:  | | : :|   :   |: : |:| :| : ||   :   | :|    |||:| |
Db     172 KFSWSSRGVLLALRNAGKWWQDGKIVEVEGKDLMKMAKPYFI--YPGYAFVAYPNRDSTI 229

Qy     785 YGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRPT------YK 838
           | : | |  || |: | | ||:|| : : ||   || |        ||:| :      :|
Db     230 YKERYNI-PEAQTVIRGTLRYQGFPQFIKTLVDIGFLD--------DTARESLSKQTPWK 280

Qy     839 GFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGL-HE 897
               |::   |:   ||:            | :|      ::|   : :  ::::||  :
Db     281 EATKEIVGAASSSQADLE------------AAILSKATFESAEDQKRILSGLRWIGLFSD 328

Qy     898 ETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEFG 957
           ||  |:|  :| | :|  :||:| :   |:|:|:| |: |:|: ||   |   :||:|:|
Db     329 ETITPRG--NPLDTLCATLEQKMQFEEGERDLVMLQHKFEIEHADGS-RETRTSTLVEYG 385

Qy     958 KVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPAL-EILESSGIKL 1016
              : :  :||| |||:| |:    :|  ::  |||: |:  :|  | : |: |  || :
Db     386 ---DPKGYSAMAKTVGVPCAVAVKQVLSGQISGKGVLAPMSTDITEPLMKELHEKYGITM 442

Qy    1017 VEK 1019
           :||
Db     443 IEK 445


RESULT  15
Q9P4R4
ID   Q9P4R4       PRELIMINARY;       PRT;    450 AA.
AC   Q9P4R4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SACCHAROPINE REDUCTASE.
GN   LYS3.
OS   Magnaporthe grisea (Rice blast fungus) (Pyricularia grisea).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
OX   NCBI_TaxID=148305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=4091-5-8;
RX   MEDLINE=20235668; PubMed=10771443;
RA   Johansson E., Steffens J.J., Emptage M., Lindqvist Y., Schneider G.;
RT   "Cloning, expression, purification and crystallization of saccharopine
RT   reductase from Magnaporthe grisea.";
RL   Acta Crystallogr. D Biol. Crystallogr. 56:662-664(2000).
DR   EMBL; AF144424; AAF91081.1; -.
SQ   SEQUENCE    450 AA;  49059 MW;  699732A79C9E5C00 CRC64;

  Query Match             13.6%;  Score 717;  DB 3;  Length 450;
  Best Local Similarity   35.6%;  Pred. No. 4.6e-41;
  Matches  170;  Conservative   90;  Mismatches  181;  Indels   36;  Gaps   10;

Qy     546 ILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGIENTTA 605
           :|:||:| | ||  : |                  | | ||    : |::   |::::|
Db       6 VLMLGSGFVTRPTLDVLTD--------------SGIKVTVACRTLESAKKLSAGVQHSTP 51

Qy     606 TQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQA 665
             ||| |  :|   |:: ::||||:| :||| :    |  |||:|| |||  :|  | ||
Db      52 ISLDVNDDAALDAEVAKHDLVISLIPYTFHATVIKSAIRQKKHVVTTSYVSPAMMELDQA 111

Qy     666 AKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYK 725
           |||||:|:: |:||||||||| ::| |:  ||  |||| | |||||||:| :::||| ||
Db     112 AKDAGITVMNEIGLDPGIDHLYAIKTIEGVHAAGGKIKTFLSYCGGLPAPESSDNPLGYK 171

Qy     726 FSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSLIY 785
           |||:  | | : :| | :   |:  :| |  |  :||   :   | ||    |||:|  |
Db     172 FSWSSRGVLLALRNAASFYKDGKVTNVAGPELMATAKPYFI--YPGFAFVAYPNRDSTPY 229

Qy     786 GDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELL 845
            : | |  ||  | | | ||:|| : :  |   ||        | |  :|    || | :
Db     230 KERYQI-PEADNIVRGTLRYQGFPQFIKVLVDIGF--------LSDEEQP----FLKEAI 276

Qy     846 NNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHEETQI-PKG 904
                  |   ::||   : |::: ::     :: |   : |  :|:||:  : :| |:|
Db     277 PWKEA--TQKIVKASSASEQDIVSTIVSNATFESTEEQKRIVAGLKWLGIFSDKKITPRG 334

Qy     905 CSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEFGKVENGRS 964
             :  | :|  :|::| :   |:|:|:| |: |:|  ||   |   ::| |:|
Db     335 --NALDTLCATLEEKMQFEEGERDLVMLQHKFEIENKDGS-RETRTSSLCEYGAPIGSGG 391

Qy     965 TTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPAL-EILESSGIKLVEKV 1020
            :|||  ||:| |:    :|   :  :||: |:  :|  | : |: |  ||:  |||
Db     392 YSAMAKFVGVPCAVAVKFVLDGTISDRGVLAPMNSKINDPLMKELKEKYGIECKEKV 448


Search completed: May 31, 2002, 15:51:48
Job time: 245 sec
```

```
                          GenCore version 4.5
                 Copyright (c) 1993 - 2000  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          June  1, 2002, 00:12:00 ; Search time 72.31 Seconds
                                           (without alignments)
                                           11091.044 Million cell updates/sec

Title:           US-09-049-304A-120
Perfect score:   3265
Sequence:        1 ATTGTGCCCGCCTTCTGCTA..........TCAAATAATGCCGATCAGTT 3265

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters:      767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
               1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
               2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
               3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
               4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
               5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
               6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.

                              SUMMARIES

                   %
Result           Query
  No.    Score   Match Length DB  ID                          Description
-------------------------------------------------------------------------------
    1    105.2    3.2    323  1  US-08-474-633A-103          Sequence 103, App
    2     92.8    2.8    372  1  US-08-474-633A-102          Sequence 102, App
c   3     45.4    1.4   7218  1  US-08-232-463-14            Sequence 14, Appl
    4     37.2    1.1   2659  1  US-08-007-775-3             Sequence 3, Appli
c   5       37    1.1   3414  1  US-08-764-100-17            Sequence 17, Appl
    6       37    1.1   3414  1  US-08-764-100-21            Sequence 21, Appl
c   7       37    1.1   4970  1  US-08-764-100-14            Sequence 14, Appl
    8       37    1.1   4970  1  US-08-764-100-20            Sequence 20, Appl
    9     36.4    1.1   5661  4  US-08-938-105-2             Sequence 2, Appli
c  10     36.2    1.1    198  5  PCT-US95-10668-1            Sequence 1, Appli
c  11     36.2    1.1    198  5  PCT-US95-10668-2            Sequence 2, Appli
c  12     36.2    1.1    198  5  PCT-US95-10668-3            Sequence 3, Appli
c  13     36.2    1.1    198  5  PCT-US95-10668-4            Sequence 4, Appli
   14       36    1.1   2105  4  US-09-088-425-3             Sequence 3, Appli
   15       36    1.1   2244  4  US-09-272-414-1             Sequence 1, Appli
c  16     35.8    1.1   8010  4  US-09-521-526-2             Sequence 2, Appli
c  17     35.8    1.1   8010  5  PCT-US95-11859-2            Sequence 2, Appli
c  18     35.4    1.1   2700  3  US-09-022-875-1             Sequence 1, Appli
   19     34.8    1.1   2671  2  US-08-666-367B-1            Sequence 1, Appli
   20     34.8    1.1   2671  4  US-09-143-438-1             Sequence 1, Appli
c  21     34.6    1.1   3489  2  US-08-728-323A-1            Sequence 1, Appli
   22     34.6    1.1  32207  2  US-08-770-379-20            Sequence 20, Appl
   23     34.6    1.1  32207  4  US-08-757-669A-20           Sequence 20, Appl
   24     34.6    1.1  32207  4  US-09-230-371A-20           Sequence 20, Appl
   25     34.4    1.1    558  2  US-08-732-791-2             Sequence 2, Appli
   26     34.4    1.1    558  3  US-09-205-049-2             Sequence 2, Appli
   27     34.4    1.1   1891  4  US-08-973-462-3             Sequence 3, Appli
```

```
   28     34.4    1.1   4376  1  US-08-119-125A-1            Sequence 1, Appli
   29     34.4    1.1   6744  1  US-08-119-125A-2            Sequence 2, Appli
c  30     34.2    1.0  15328  2  US-08-888-497-33            Sequence 33, Appl
c  31     34.2    1.0  15328  4  US-09-362-230-33            Sequence 33, Appl
c  32     34.2    1.0  15328  5  PCT-US94-07926-33           Sequence 33, Appl
c  33       34    1.0   3642  3  US-08-946-026-16            Sequence 16, Appl
   34     33.8    1.0   2080  1  US-08-179-738-9             Sequence 9, Appli
   35     33.8    1.0   2080  2  US-08-628-145-9             Sequence 9, Appli
   36     33.8    1.0   2257  1  US-08-171-718-15            Sequence 15, Appl
   37     33.8    1.0   2257  3  US-08-478-087-15            Sequence 15, Appl
c  38     33.4    1.0   9785  4  US-09-479-128-1             Sequence 1, Appli
   39     33.2    1.0   1150  1  US-08-161-406-1             Sequence 1, Appli
   40     33.2    1.0   6038  4  US-09-305-639-4             Sequence 4, Appli
   41     33.2    1.0   7622  4  US-09-305-639-1             Sequence 1, Appli
   42     33.2    1.0   8532  1  US-08-452-655B-1            Sequence 1, Appli
   43     33.2    1.0   8532  3  US-08-450-582-1             Sequence 1, Appli
c  44     33.2    1.0   8835  3  US-08-884-324-10            Sequence 10, Appl
   45     33.2    1.0   9606  1  US-07-741-940-1             Sequence 1, Appli

                              ALIGNMENTS

RESULT   1
US-08-474-633A-103
; Sequence 103, Application US/08474633A
; Patent No. 5773691
;  GENERAL INFORMATION:
;    APPLICANT:  E. I. DU PONT DE NEMOURS AND
;    APPLICANT:  COMPANY
;    TITLE OF INVENTION:  CHIMERIC GENES AND
;    TITLE OF INVENTION:  METHODS FOR INCREASING
;    TITLE OF INVENTION:  INCREASING THE LYSINE
;    TITLE OF INVENTION:  AND THREONINE CONTENT
;    TITLE OF INVENTION:  OF THE SEEDS OF PLANTS
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS
;      ADDRESSEE:  AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  MICROSOFT WORD VERSION 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,633A
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BARBARA C. SIEGELL
;      REGISTRATION NUMBER:  30,684
;      REFERENCE/DOCKET NUMBER:  BB-1037-C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-992-4931
;      TELEFAX:  302-773-0164
;      TELEX:  835420
;  INFORMATION FOR SEQ ID NO:  103:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  323 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-474-633A-103

  Query Match             3.2%;  Score 105.2;  DB 1;  Length 323;
  Best Local Similarity   65.1%;  Pred. No. 5.3e-24;
```

```
  Matches  155;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy   2844 AAGCACCAAGCGACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCACCACTGCC 2903
          ||||||   ||||| ||  |||| |||||| |  |  | |||||     | || || ||
Db      3 AAGCACACTGCGACTCTTTTGGAATTCGGGGACATCAAGAATGGACAAACAACAACCGCT 62

Qy   2904 ATGGCGCTGACCGTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAG 2963
          |||||   ||| || || || || ||||| || || || ||||| ||  |  |  | |||
Db     63 ATGGCCAAGACTGTTGGGATCCCTGCAGCCATTGGAGCTCTGCTGTTAATTGAAGACAAG 122

Qy   2964 GTCCAGACGAAAGGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAGCATTGGAG 3023
           || |||| | ||||||  | ||||||||  || | ||  | ||  | ||||| |||||
Db    123 ATCAAGACAAGAGGAGTCTTAAGGCCTCTCGAAGCAGAGGTGTATTTGCCAGCTTTGGAT 182

Qy   3024 ATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGAAAGTTCCCTG 3081
          || ||| |  | |  || || |||||| | ||||| |  ||    | ||||  | |||
Db    183 ATATTGCAAGCATATGGTATAAAGCTGATGGAGAAGGCAGAATGATCAAAGAACTCTG 240


RESULT   2
US-08-474-633A-102
; Sequence 102, Application US/08474633A
; Patent No. 5773691
;  GENERAL INFORMATION:
;    APPLICANT:  E. I. DU PONT DE NEMOURS AND
;    APPLICANT:  COMPANY
;    TITLE OF INVENTION:  CHIMERIC GENES AND
;    TITLE OF INVENTION:  METHODS FOR INCREASING
;    TITLE OF INVENTION:  INCREASING THE LYSINE
;    TITLE OF INVENTION:  AND THREONINE CONTENT
;    TITLE OF INVENTION:  OF THE SEEDS OF PLANTS
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS
;      ADDRESSEE:  AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  MICROSOFT WORD VERSION 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,633A
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BARBARA C. SIEGELL
;      REGISTRATION NUMBER:  30,684
;      REFERENCE/DOCKET NUMBER:  BB-1037-C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-992-4931
;      TELEFAX:  302-773-0164
;      TELEX:  835420
;  INFORMATION FOR SEQ ID NO:  102:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  372 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-474-633A-102

  Query Match              2.8%;  Score 92.8;  DB 1;  Length 372;
  Best Local Similarity   61.1%;  Pred. No. 6.3e-20;
  Matches  179;  Conservative    0;  Mismatches   93;  Indels   21;  Gaps    1;

Qy   1639 TTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTGAGTTTCTGGCATCTTACC 1698
          |||| |||||||| |||||| | || | ||| ||||||||||| || || || ||
```

```
Db     70 TTTTGATTCTTGGTGCTGGACGTGTGTNTCGCCCAGCTGCTGATTTCCTAGCTTCAGTTA 129

Qy   1699 CAGACATATGTACCTAT---------------------GGTGTTGATGACCATGATGCAG 1737
           |  ||| |   |  |                      || |  ||     | ||   |
Db    130 GAACCATTTCGTCACAGCAATGGTACAAAACATATTTCGGAGCAGACTCTGAAGAGAAAA 189

Qy   1738 ATCAAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATG 1797
             |  ||||||| || || || ||| ||||||  || |||||  ||||||| ||||| |
Db    190 CAGATGTTCATGTGATTGTCGCGTCTCTGTATCTTAAGGATGCCAAAGAGACGGTTGAAG 249

Qy   1798 GTATTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATC 1857
          |||||  | ||  |   ||    | ||| |||||  |||||| || ||| ||     |
Db    250 GTATTTCAGATGTAGAAGCAGTTCGGCTAGATGTATCTGATAGTGAAAGTCTCCTTAAGT 309

Qy   1858 TTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCT 1910
           ||||||||||||||| |||||  | || ||  | ||||| |||| |||||||
Db    310 ATGTTTCTCAGGTTGATGTTGTCCTAAGTTTATTACCTGCAAGTTGTCATGCT 362


RESULT   3
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match              1.4%;  Score 45.4;  DB 1;  Length 7218;
  Best Local Similarity    7.6%;  Pred. No. 0.0016;
  Matches   34;  Conservative  217;  Mismatches  198;  Indels    0;  Gaps    0;
```

```
Qy    2735 GATTTGCCAGCGAATGGAACAGAGGATGGCCTATGGCCACAATGAGCAAGACATGGTACT 2794
           || ||     |::: :::: :::::: ::   : ::  : :: ::: :::: : :: :
Db    1445 GAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1386

Qy    2795 GCTCCACCACGAAGTCGAGGTGGAATACCCGGACGGGCAACCCGCCGAAAAGCACCAAGC 2854
           :    :  : ::::  :::: :::: :   ::: ::: ::   :  :::::: :  :::
Db    1385 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1326

Qy    2855 GACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCACCACTGCCATGGCGCTGAC 2914
           :: :  :  ::::   :::::::  ::::: :: :::   :  :  :  : :: :  ::
Db    1325 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1266

Qy    2915 CGTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAGGTCCAGACGAA 2974
            :  :: :    :: :: :: :::::    :  :  :  :::::: ::::   ::: :::
Db    1265 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1206

Qy    2975 AGGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAGCATTGGAGATCTTGGAGTC 3034
           ::::: ::  :::     : ::  :::::   : :    :: :  :::::    ::::
Db    1205 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1146

Qy    3035 GTCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGAAAGTTCCCTGATACACAGATAAA 3094
           :  ::: :  :::  ::  ::::::: :::::   :::::      :: : : ::: :::
Db    1145 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1086

Qy    3095 GATAGTATGATATAGCAGGGCACATGTATCTTTTGTATTAACTCCGTTCTGGAATATATA 3154
           :: :: : :: : :: :::         ||  | |   |    ||   || | |  ||
Db    1085 RRRRRRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACCTGCAGCCAAGCTCGGAATTAAT 1026

Qy    3155 TTTGTGAACTAAAATGTGACAAATAAAAA 3183
           | ||||| |  |      || ||  ||||
Db    1025 TCTGTGAGCGTATGGCAAACGAAGGAAAA 997

RESULT    4
US-08-007-775-3
; Sequence 3, Application US/08007775
; Patent No. 5340733
;  GENERAL INFORMATION:
;    APPLICANT:  Takashi UENO et al.
;    TITLE OF INVENTION:  MboI RESTRICTION-MODIFICATION GENES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 500 kb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/007,775
;      FILING DATE:  19930122
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2659 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:  Moraxella bovis
;      STRAIN:  10900
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;      UNITS:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:  12
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:  /note= "inosine"
;    PUBLICATION INFORMATION:
;      AUTHORS:
;      TITLE:
;      JOURNAL:
;      VOLUME:
;      ISSUE:
;      PAGES:
;      DATE:
;      DOCUMENT NUMBER:
;      FILING DATE:
;      PUBLICATION DATE:
;      RELEVANT RESIDUES IN SEQ ID NO:
US-08-007-775-3

  Query Match              1.1%;  Score 37.2;  DB 1;  Length 2659;
  Best Local Similarity   51.4%;  Pred. No. 0.34;
  Matches  110;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy    1679 TGAGTTTCTGGCATCTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGA 1738
           ||||    |   | || |||    |||  || || || ||  || |  | |  | | | |
Db     766 TGAGAGAATACAAGCTAACCGCACCATTAGTGCCAATAGTAATGGTCGCAAAAAGGTAAA 825

Qy    1739 TCAAATTCATGTTATCGTGGCATCTTTGTATCAAA-AAGATGCAGAAGAGACAGTTGATG 1797
           | |||||   || |     | |  ||   || |||  ||     || ||     |  ||
Db     826 TGAAATTATAGTCAGTAACGGAGTTTAACATGAAATTAGCATTTGATGATTTTTTAAATA 885

Qy    1798 GTATTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATC 1857
           ||||   | | || | | ||||   |  | || || ||||| |||  |   |   | |
Db     886 GTATGTCAGAAACCAATACTACTTTGGATTATTTTACTGATTTTGATAAAGTAAAAAAGA 945

Qy    1858 TTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCT 1891
            |||  ||||  ||||| ||    | | |  |||
Db     946 ATGTCGCTCAAATTGAGATTCACTTAAACCAGCT 979

RESULT    5
US-08-764-100-17/c
; Sequence 17, Application US/08764100
; Patent No. 5773700
;  GENERAL INFORMATION:
;    APPLICANT:  van Grinsven J., Martinus Q.
;    APPLICANT:  De Haan, Petrus T.
;    APPLICANT:  Gielen L., Johannes J.
;    APPLICANT:  Peters, Dirk
;    APPLICANT:  Goldbach, Robert W.
```

```
;    TITLE OF INVENTION:  Improvements in or Relating to Organic
;    TITLE OF INVENTION:  Compounds
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sandoz Agro, Inc
;      STREET:  975 California Avenue
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,100
;      FILING DATE:  06-DEC-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/214,064
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/032,235
;      FILING DATE:  17-MAR-1993
;      APPLICATION NUMBER:  GB 9206016.9
;      FILING DATE:  19-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5773700ris, Allen E.
;      REGISTRATION NUMBER:  34,490
;      REFERENCE/DOCKET NUMBER:  137-1061
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 354-3592
;      TELEFAX:  (415) 857-1125
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3414 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-764-100-17

  Query Match             1.1%;  Score 37;  DB 1;  Length 3414;
  Best Local Similarity   53.0%;  Pred. No. 0.49;
  Matches   79;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy   1694 TTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTAT 1753
          | | ||||| || | |    | | |  | |||||| |  || |||||| |   |||  ||
Db   3286 TCATCCAGAAATTTATGATGAAGCTTATTATGACCGTTCTGTAGATCATAAAAATGAAAT 3227

Qy   1754 CGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAAC 1813
            | |   |||||  | |||         | |  |||| | |     ||  |   |||
Db   3226 TCTAGATACTTTGGCTGAAATGCTCCAGAATGCAACAGGTAAAACCCTTAGACCAACACG 3167

Qy   1814 TGCTACCCAGCTTGATGTTGCTGATATTG 1842
           | ||| ||   || | ||||  ||| ||
Db   3166 AGATACTCAAACTGTTCTTGCCAATAATG 3138

RESULT    6
US-08-764-100-21
; Sequence 21, Application US/08764100
; Patent No. 5773700
;  GENERAL INFORMATION:
;    APPLICANT:  van Grinsven J., Martinus Q.
;    APPLICANT:  De Haan, Petrus T.
;    APPLICANT:  Gielen L., Johannes J.
;    APPLICANT:  Peters, Dirk
;    APPLICANT:  Goldbach, Robert W.
;    TITLE OF INVENTION:  Improvements in or Relating to Organic
;    TITLE OF INVENTION:  Compounds
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sandoz Agro, Inc
;      STREET:  975 California Avenue
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,100
;      FILING DATE:  06-DEC-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/214,064
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/032,235
;      FILING DATE:  17-MAR-1993
;      APPLICATION NUMBER:  GB 9206016.9
;      FILING DATE:  19-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5773700ris, Allen E.
;      REGISTRATION NUMBER:  34,490
;      REFERENCE/DOCKET NUMBER:  137-1061
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 354-3592
;      TELEFAX:  (415) 857-1125
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3414 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-764-100-21

  Query Match             1.1%;  Score 37;  DB 1;  Length 3414;
  Best Local Similarity   53.0%;  Pred. No. 0.49;
  Matches   79;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy   1694 TTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTAT 1753
          | | ||||| || | |    | | |  | |||||| |  || |||||| |   |||  ||
Db    129 TCATCCAGAAATTTATGATGAAGCTTATTATGACCGTTCTGTAGATCATAAAAATGAAAT 188

Qy   1754 CGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAAC 1813
            | |   |||||  | |||         | |  |||| | |     ||  |   |||
Db    189 TCTAGATACTTTGGCTGAAATGCTCCAGAATGCAACAGGTAAAACCCTTAGACCAACACG 248

Qy   1814 TGCTACCCAGCTTGATGTTGCTGATATTG 1842
           | ||| ||   || | ||||  ||| ||
Db    249 AGATACTCAAACTGTTCTTGCCAATAATG 277

RESULT    7
US-08-764-100-14/c
; Sequence 14, Application US/08764100
; Patent No. 5773700
;  GENERAL INFORMATION:
;    APPLICANT:  van Grinsven J., Martinus Q.
;    APPLICANT:  De Haan, Petrus T.
;    APPLICANT:  Gielen L., Johannes J.
;    APPLICANT:  Peters, Dirk
;    APPLICANT:  Goldbach, Robert W.
;    TITLE OF INVENTION:  Improvements in or Relating to Organic
;    TITLE OF INVENTION:  Compounds
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sandoz Agro, Inc
;      STREET:  975 California Avenue
;      CITY:  Palo Alto
;      STATE:  CA
```

```
;       COUNTRY:  USA
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/764,100
;       FILING DATE:  06-DEC-1996
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/214,064
;       FILING DATE:
;       APPLICATION NUMBER:  US 08/032,235
;       FILING DATE:  17-MAR-1993
;       APPLICATION NUMBER:  GB 9206016.9
;       FILING DATE:  19-MAR-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 5773700ris, Allen E.
;       REGISTRATION NUMBER:  34,490
;       REFERENCE/DOCKET NUMBER:  137-1061
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 354-3592
;       TELEFAX:  (415) 857-1125
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4970 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-764-100-14

  Query Match             1.1%;  Score 37;  DB 1;  Length 4970;
  Best Local Similarity   53.0%;  Pred. No. 0.65;
  Matches   79;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    1694 TTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTAT 1753
           | | ||||| || | |    | | |  | |||||| |  || |||||| |   |||  ||
Db    4756 TCATCCAGAAATTTATGATGAAGCTTATTATGACCGTTCTGTAGATCATAAAAATGAAAT 4697

Qy    1754 CGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAAC 1813
             | |   |||||  | |||         | |  |||| | |     ||  |   |||
Db    4696 TCTAGATACTTTGGCTGAAATGCTCCAGAATGCAACAGGTAAAACCCTTAGACCAACACG 4637

Qy    1814 TGCTACCCAGCTTGATGTTGCTGATATTG 1842
            | ||| ||   || | ||||  ||| ||
Db    4636 AGATACTCAAACTGTTCTTGCCAATAATG 4608

RESULT   8
US-08-764-100-20
; Sequence 20, Application US/08764100
; Patent No. 5773700
;  GENERAL INFORMATION:
;    APPLICANT:  van Grinsven J., Martinus Q.
;    APPLICANT:  De Haan, Petrus T.
;    APPLICANT:  Gielen L., Johannes J.
;    APPLICANT:  Peters, Dirk
;    APPLICANT:  Goldbach, Robert W.
;    TITLE OF INVENTION:  Improvements in or Relating to Organic
;    TITLE OF INVENTION:  Compounds
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sandoz Agro, Inc
;      STREET:  975 California Avenue
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/764,100
;       FILING DATE:  06-DEC-1996
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/214,064
;       FILING DATE:
;       APPLICATION NUMBER:  US 08/032,235
;       FILING DATE:  17-MAR-1993
;       APPLICATION NUMBER:  GB 9206016.9
;       FILING DATE:  19-MAR-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 5773700ris, Allen E.
;       REGISTRATION NUMBER:  34,490
;       REFERENCE/DOCKET NUMBER:  137-1061
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 354-3592
;       TELEFAX:  (415) 857-1125
;   INFORMATION FOR SEQ ID NO:  20:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4970 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-764-100-20

  Query Match             1.1%;  Score 37;  DB 1;  Length 4970;
  Best Local Similarity   53.0%;  Pred. No. 0.65;
  Matches   79;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    1694 TTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTAT 1753
           | | ||||| || | |    | | |  | |||||| |  || |||||| |   |||  ||
Db     215 TCATCCAGAAATTTATGATGAAGCTTATTATGACCGTTCTGTAGATCATAAAAATGAAAT 274

Qy    1754 CGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAAC 1813
             | |   |||||  | |||         | |  |||| | |     ||  |   |||
Db     275 TCTAGATACTTTGGCTGAAATGCTCCAGAATGCAACAGGTAAAACCCTTAGACCAACACG 334

Qy    1814 TGCTACCCAGCTTGATGTTGCTGATATTG 1842
            | ||| ||   || | ||||  ||| ||
Db     335 AGATACTCAAACTGTTCTTGCCAATAATG 363

RESULT   9
US-08-938-105-2
; Sequence 2, Application US/08938105
; Patent No. 6353151
;  GENERAL INFORMATION:
;    APPLICANT:  Leinwand, Leslie A.
;    APPLICANT:  Vikstrom, Karen L.
;    TITLE OF INVENTION:  TRANSGENIC MODEL FOR HEART FAILURE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross P.C.
;      STREET:  1700 Lincoln St., Suite 3500
;      CITY:  Denver
;      STATE:  CO
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/938,105
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Crook, Wannell M.
;       REGISTRATION NUMBER:  31,071
;       REFERENCE/DOCKET NUMBER:  3595-4
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (303) 863-9700
;       TELEFAX:  (303) 863-0223
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5661 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..5661
US-08-938-105-2

  Query Match              1.1%;  Score 36.4;  DB 4;  Length 5661;
  Best Local Similarity   51.9%;  Pred. No. 1.1;
  Matches   82;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy   2720 GAGCCCATTTGATGTGATTTGCCAGCGAATGGAACAGAGGATGGCCTATGGCCACAATGA 2779
          || ||     || | ||| |   | |  |  ||||||  |   |     ||  | |||| 
Db   4356 GAACCTCCAGGAGGAGATCTCAGACCTGACTGAACAGCTGGGAGAAGGGGGTAAAAATGT 4415

Qy   2780 GCAAGACATGGTACTGCTCCACCACGAAGTCGAGGTGGAATACCCGGACGGGCAACCCGC 2839
          ||| ||  |||    | ||| | |  |  | ||||||||  | | |||   |||  | ||
Db   4416 GCACGAGCTGGAGAAGATCCGCAAACAGCTGGAGGTGGAGAAGCTGGAACTGCAGTCAGC 4475

Qy   2840 CGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGAAGG 2877
          |    || |    | | |    ||||||   | | |||
Db   4476 CCTGGAGGAGGCTGAGGCCTCCCTGGAGCATGAGGAGG 4513


RESULT  10
PCT-US95-10668-1/c
; Sequence 1, Application PC/TUS9510668
;  GENERAL INFORMATION:
;    APPLICANT:  James Eberwine
;    TITLE OF INVENTION:  A Method of Sequencing Proteins by
;    TITLE OF INVENTION:  Epitope Ordering and Protein
;    TITLE OF INVENTION:  Restriction Mapping
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jane Massey Licata, Esq.
;      STREET:  210 Lake Drive East, Suite 201
;      CITY:  Cherry Hill
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10668
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/294,133
;      FILING DATE:  August 22, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  PENN-0137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 779-2400
;      TELEFAX:  (609) 779-8488
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
```

```
;       LENGTH:  198
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
;     ANTI-SENSE:  No
PCT-US95-10668-1

  Query Match              1.1%;  Score 36.2;  DB 5;  Length 198;
  Best Local Similarity   50.9%;  Pred. No. 0.096;
  Matches   86;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy   1715 TGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATCTTTGTATCAAAA 1774
          || ||||| ||    || ||  | |    ||  |||| | || |    | |   |
Db    170 TGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 111

Qy   1775 AGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCAGCTTGATGTTGC 1834
           | ||  |  |     |||| || | |||    |     || |       ||| |||||
Db    110 TGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 51

Qy   1835 TGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATT 1883
          || | |||| | ||||   | | |||||  |   ||||  |||||  ||
Db     50 TGTTGTTGGGATCCTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2


RESULT  11
PCT-US95-10668-2/c
; Sequence 2, Application PC/TUS9510668
;  GENERAL INFORMATION:
;    APPLICANT:  James Eberwine
;    TITLE OF INVENTION:  A Method of Sequencing Proteins by
;    TITLE OF INVENTION:  Epitope Ordering and Protein
;    TITLE OF INVENTION:  Restriction Mapping
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jane Massey Licata, Esq.
;      STREET:  210 Lake Drive East, Suite 201
;      CITY:  Cherry Hill
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10668
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/294,133
;      FILING DATE:  August 22, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  PENN-0137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 779-2400
;      TELEFAX:  (609) 779-8488
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  198
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    ANTI-SENSE:  No
PCT-US95-10668-2

  Query Match              1.1%;  Score 36.2;  DB 5;  Length 198;
  Best Local Similarity   50.9%;  Pred. No. 0.096;
  Matches   86;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;
```

```
Qy   1715 TGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATCTTTGTATCAAAA 1774
          || ||||| ||    || ||  | |    ||  |||| | || |    | |   |
Db    170 TGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 111

Qy   1775 AGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCAGCTTGATGTTGC 1834
           | ||  |  |     |||| || | |||    |     || |       ||| |||||
Db    110 TGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 51

Qy   1835 TGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATT 1883
          || | |||| | ||||   | | |||||  |   ||||  |||||  ||
Db     50 TGTTGTTGGGATCCTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2


RESULT  12
PCT-US95-10668-3/c
; Sequence 3, Application PC/TUS9510668
;  GENERAL INFORMATION:
;    APPLICANT:  James Eberwine
;    TITLE OF INVENTION:  A Method of Sequencing Proteins by
;    TITLE OF INVENTION:  Epitope Ordering and Protein
;    TITLE OF INVENTION:  Restriction Mapping
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jane Massey Licata, Esq.
;      STREET:  210 Lake Drive East, Suite 201
;      CITY:  Cherry Hill
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10668
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/294,133
;      FILING DATE:  August 22, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  PENN-0137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 779-2400
;      TELEFAX:  (609) 779-8488
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  198
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    ANTI-SENSE:  No
PCT-US95-10668-3


  Query Match             1.1%;  Score 36.2;  DB 5;  Length 198;
  Best Local Similarity   50.9%;  Pred. No. 0.096;
  Matches   86;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy   1715 TGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATCTTTGTATCAAAA 1774
          || ||||| ||    || ||  | |    ||  |||| | || |    | |   |
Db    170 TGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 111

Qy   1775 AGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCAGCTTGATGTTGC 1834
           | ||  |  |     |||| || | |||    |     || |       ||| |||||
Db    110 TGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 51

Qy   1835 TGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATT 1883
          || | ||||||  |||   | | |||||  |   ||||  |||||  ||
Db     50 TGTTGTTGGAATTCTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2

RESULT  13
PCT-US95-10668-4/c
; Sequence 4, Application PC/TUS9510668
;  GENERAL INFORMATION:
;    APPLICANT:  James Eberwine
;    TITLE OF INVENTION:  A Method of Sequencing Proteins by
;    TITLE OF INVENTION:  Epitope Ordering and Protein
;    TITLE OF INVENTION:  Restriction Mapping
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jane Massey Licata, Esq.
;      STREET:  210 Lake Drive East, Suite 201
;      CITY:  Cherry Hill
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08002
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10668
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/294,133
;      FILING DATE:  August 22, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  PENN-0137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 779-2400
;      TELEFAX:  (609) 779-8488
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  198
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    ANTI-SENSE:  No
PCT-US95-10668-4


  Query Match             1.1%;  Score 36.2;  DB 5;  Length 198;
  Best Local Similarity   50.9%;  Pred. No. 0.096;
  Matches   86;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy   1715 TGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATCTTTGTATCAAAA 1774
          || ||||| ||    || ||  | |    ||  |||| | || |    | |   |
Db    170 TGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 111

Qy   1775 AGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCAGCTTGATGTTGC 1834
           | ||  |  |     |||| || | |||    |     || |       ||| |||||
Db    110 TGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGT 51

Qy   1835 TGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATT 1883
          || | ||||||  |||   | | |||||  |   ||||  |||||  ||
Db     50 TGTTGTTGGAATTCTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2

RESULT  14
US-09-088-425-3
; Sequence 3, Application US/09088425
; Patent No. 6171843
;  GENERAL INFORMATION:
;    APPLICANT:  BANDMAN, OLGA
;    APPLICANT:  LAL, PREETI
;    APPLICANT:  CORLEY, NEIL C.
```

```
;    APPLICANT:  PATTERSON, CHANDRA
;    APPLICANT:  BAUGHN, MARIAH R.
;    TITLE OF INVENTION:  HUMAN ISOMERASE-LIKE PROTEINS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/088,425
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0529 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2105 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BRAINON01
;      CLONE:  2291164
US-09-088-425-3

  Query Match             1.1%;  Score 36;  DB 4;  Length 2105;
  Best Local Similarity   53.6%;  Pred. No. 0.7;
  Matches   75;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy   1703 CATATGTACCTATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATC 1762
          || | |||   |||   |   | |  ||||| || |||  | |  ||||   || | |  
Db    890 CAAAAGTAGTCATGACTTGCTTAAGGATGATCCACATCTCAGTTCTGTTCCAGTTGTAGA 949

Qy   1763 TTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCA 1822
               |  ||   ||||||  |||   |||||||| |   ||| ||  ||  |||    ||
Db    950 AAGTGAAAAAGGTGATGCACCAGATTTAGTTGATGATGGAGAAGATGAAAGTGCAGAGCA 1009

Qy   1823 GCTTGATGTTGCTGATATTG 1842
             |||   |  ||||  ||
Db   1010 TGATGAATATATTGATGGTG 1029


RESULT  15
US-09-272-414-1
; Sequence 1, Application US/09272414A
; Patent No. 6238885
; GENERAL INFORMATION:
;  APPLICANT: Wallis, Nicola G.
;  TITLE OF INVENTION: Histidine Kinase
;  FILE REFERENCE: GM10202
;  CURRENT APPLICATION NUMBER: US/09/272,414A
;  CURRENT FILING DATE: 1999-03-19
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;    LENGTH: 2244
;    TYPE: DNA
;    ORGANISM: Staphylococcus aureus
US-09-272-414-1

  Query Match             1.1%;  Score 36;  DB 4;  Length 2244;
  Best Local Similarity   51.2%;  Pred. No. 0.74;
  Matches   84;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy   1464 GAAGTAGGAGCAGATGATACTGCCACATTGGATAAAATTATTGATTCCTTGACTTCTTTA 1523
          |||| || |  | |  ||| |   |||||   | ||| |  || |   ||||    | ||
Db    279 gaagaagcaataaaatatagtcaaacattaattgaaaatcctggtgggttgatgattata 338

Qy   1524 GCTAATGAACATGGTGGAGATCACGATGCCGGGCAAGAAATTGAATTAGCTCTGAAGATA 1583
            |||| ||||| ||  |      | | |        | ||| ||        ||  ||  
Db    339 aataataaacatcgtcaatcaacggcttcactttctaatattaaaaagcaaatgttgaat 398

Qy   1584 GGAAAAGTCAATGAGTATGAAACTGACGTCACAATTGATAAAGG 1627
          | |  ||||||  |  |  |   |||||     ||||||||||
Db    399 gaagtagtcaacaacgaccattttgacgatgtgtttgataaagg 442


Search completed: June  1, 2002, 02:05:08
Job time: 6788 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         June  1, 2002, 00:15:00 ; Search time 346.29 Seconds
                                           (without alignments)
                                           16187.951 Million cell updates/sec

Title:          US-09-049-304A-120
Perfect score:  3265
Sequence:       1 ATTGTGCCCGCCTTCTGCTA..........TCAAATAATGCCGATCAGTT 3265

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters:      3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      N_Geneseq_032802:*
             1:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
             2:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
             3:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
             4:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
             5:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
             6:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
             7:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
             8:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
             9:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
            10:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
            11:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
            12:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
            13:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
            14:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
            15:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
            16:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
            17:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
            18:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
            19:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
            20:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
            21:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
            22:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
            23:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
            24:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                        %
Result                Query
  No.     Score       Match Length DB  ID              Description
--------------------------------------------------------------------------
    1      3263       99.9    3265 20  AAV99563        Maize lysine ketog
    2      1908       58.4    1908 20  AAV99564        Maize lysine ketog
    3    1233.2       37.8    3195 21  AAC42953        Arabidopsis thalia
    4       907       27.8    2582 20  AAV99562        Soybean lysine ket
    5     522.6       16.0     720 20  AAV99565        Rice lysine ketogl
    6       515       15.8    1449 20  AAV99568        Arabidopsis saccha
    7     183.6        5.6     429 20  AAV99567        Wheat lysine ketog
    8     155.8        4.8    2919 23  ABL12397        Drosophila melanog
    9     155.8        4.8    3266 22  AAD08977        Drosophila melanog
c  10     110.4        3.4    6209 23  ABL12396        Drosophila melanog
   11     105.2        3.2     323 19  AAV35857        cDNA encoding prot
   12     105.2        3.2     323 20  AAV99554        Arabidopsis lysine
   13       101        3.1     308 20  AAV99566        Rice lysine ketogl
   14      97.6        3.0    1262 21  AAF13730        Aspergillus oryzae
   15      95.8        2.9    8160 20  AAV99559        Arabidopsis lysine
   16      92.8        2.8     372 19  AAV35856        cDNA encoding prot
   17      92.8        2.8     372 20  AAV99553        Arabidopsis lysine
   18      79.4        2.4     598 21  AAF07883        Fusarium venenatum
   19      49.2        1.5    4590 22  AAH24065        Yeast AOD9604-asso
   20      41.6        1.3    1830 23  AAS53006        Enterococcus faeca
c  21      41.6        1.3    2717 20  AAX13034        Enterococcus faeca
   22      41.4        1.3    2331 22  AAH16467        Human cDNA sequenc
c  23      41.4        1.3    4840 23  ABL16917        Drosophila melanog
   24      41.4        1.3   18189 23  ABL16916        Drosophila melanog
   25      41.2        1.3     556 22  ABA62873        Human foetal liver
   26      41.2        1.3     556 22  ABA30153        Probe #8619 for ge
   27      41.2        1.3     556 22  AAI17913        Probe #7846 for ge
   28      41.2        1.3     556 22  AAI42893        Probe #11579 used
c  29      40.4        1.2    2693 19  AAV71119        Exemplery caffeic
   30        40        1.2     608 24  ABI99273        Mouse ischaemic co
c  31        40        1.2    4590 22  AAH24065        Yeast AOD9604-asso
c  32      39.8        1.2    2157 23  AAS69144        DNA encoding novel
c  33      39.6        1.2     486 22  AAF75507        Polyglutamine trac
c  34      39.6        1.2    2169 21  AAA30203        Human RING finger
c  35      39.6        1.2   15016 20  AAX99560        Nucleic acid seque
c  36      39.2        1.2     531 22  AAK32668        Human bone marrow
c  37      39.2        1.2     531 22  AAI38483        Probe #7169 used t
c  38      38.8        1.2    6558 21  AAZ45602        cDNA sequence of a
c  39      38.6        1.2    3114 23  ABL26943        Drosophila melanog
   40      38.6        1.2    5114 23  ABL26942        Drosophila melanog
   41      38.6        1.2   11046 24  AAS16913        Human G-protein co
   42      38.4        1.2     222 22  ABA70505        Human foetal liver
   43      38.4        1.2     222 22  ABA37133        Probe #15599 for g
   44      38.4        1.2     222 22  AAK18749        Human brain expres
   45      38.4        1.2     222 22  AAK44685        Human bone marrow


                              ALIGNMENTS


RESULT   1
AAV99563
ID   AAV99563 standard; cDNA to mRNA; 3265 BP.
XX
AC   AAV99563;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Maize lysine ketoglutarate reductase cDNA.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; soybean; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..3071
FT                   /*tag=  a
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;    98WO-US06051.
XX
PR   27-MAR-1997;    97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
```

```
DR   P-PSDB; AAW87761.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 3; Page 184-186; 231pp; English.
XX
CC   This is the nucleotide sequence of a near full-length cDNA encoding
CC   maize lysine ketoglutarate reductase (LKR, see AAW87761), or
CC   saccharopine dehydrogenase (SDH).  It was obtained by PCR
CC   amplification of developing seed cDNA using primers (see AAV99560-61)
CC   based on Arabidopsis LKR/SDH sequences, and by RACE.  Isolated
CC   nucleic acids comprising sequences encoding all or part of plant
CC   LKRs are new.  Also claimed are: (1) a chimeric gene comprising the
CC   fragment (or a subfragment) operably linked to a seed specific
CC   regulator, where the chimeric gene reduces LKR activity in plant
CC   seeds transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence 3265 BP; 952 A; 651 C; 764 G; 897 T; 1 other;

  Query Match             99.9%;  Score 3263;  DB 20;  Length 3265;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3264;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATTGTGCCCGCCTTCTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 attgtgcccgccttctgctaggaggaggcaagaacggacctcgagtaaaccggattattg 60

Qy         61 TGCAGCCAAGCACAAGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 tgcagccaagcacaaggaggatccatcatgacgctcagtatgaggatgcaggatgcgaga 120

Qy        121 TTTCAGAAGACCTGTCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 tttcagaagacctgtcagaatgcggccttatcataggcatcaaacaacccaagctgcaga 180

Qy        181 TGATTCTTTCAGATAGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 tgattctttcagatagagcgtacgctttcttttcacacacacacaaagcccaaaaagaga 240

Qy        241 ATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 atatgccactgttagacaagatccttgaagaaagggtgtccttgtttgattatgagctaa 300

Qy        301 TTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ttgttggagatgatgggaaaagatcactagcatttgggaaatttgctggtagagctggac 360

Qy        361 TGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 tgatagatttcttacatggtctcggacagcgatatttgagccttggatactcgactccat 420

Qy        421 TTCTCTCTCTGGGACAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCA 480
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        421 ttctctctctgggacaatctcatatgtatccttctctcgctgcagccaaggctgcagtca 480

Qy        481 TTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATAGTGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ttgtcgttgcagaagagatagcaacatttggacttccatccggaatttgtccgatagtgt 540

Qy        541 TTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ttgtgttcactggagttggaaacgtctctcagggtgcgcaggagatattcaagttattgc 600

Qy        601 CCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 cccatacctttgttgatgctgagaagcttcccgaaatttttcaggccaggaatctgtcta 660

Qy        661 AGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 agcaatctcagtcgaccaagagagtatttcaactttatggttgtgttgtgacctctagag 720

Qy        721 ACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 acatagtttctcacaaggatcccaccagacaatttgacaaaggtgactattatgctcatc 780

Qy        781 CAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 cagaacactacacccctgtttttcatgaaagaattgctccatatgcatctgtcatcgtaa 840

Qy        841 ACTGTATGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 actgtatgtattgggagaagaggtttccaccattactaaatatggatcagttacagcaat 900

Qy        901 TGATGGAGACTGGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 tgatggagactggttgtcctttagtcggcgtttgtgacataacttgtgatattggaggtt 960

Qy        961 CCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ccattgaatttatcaacaagagtacatcaatagagaggcctttctttcggtatgatcctt 1020

Qy       1021 CTAAGAATTCATACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ctaagaattcataccatgatgatatggaaggtgccggagtggtctgcttggctgttgaca 1080

Qy       1081 TTCTCCCTACAGAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ttctccctacagaattctctaaagaggcctcccaacattttggaaacatactatctagac 1140

Qy       1141 TTGTTGCTAGTTTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ttgttgctagtttggcctcagtgaagcaaccggcagaacttccttcctacttgagaagag 1200

Qy       1201 CTTGCATTGCACATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 cttgcattgcacatgctggcagattaactcctttgtatgaatatatccctaggatgagaa 1260

Qy       1261 ATACTATGATAGATTTGGCACCCGCAAAAACAAATCCATTGCCTGACAAGAAGTATAGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 atactatgatagatttggcacccgcaaaaacaaatccattgcctgacaagaagtatagca 1320

Qy       1321 CCCTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ccctggtatctctcagtgggcacctatttgataagttccttataaatgaagctttggaca 1380

Qy       1381 TCATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 tcattgagacagctggaggttcatttcacttggttagatgtgaagttggacaaagcacgg 1440

Qy       1441 ATGATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 atgatatgtcatactcagagcttgaagtaggagcagatgatactgccacattggataaaa 1500

Qy       1501 TTATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCAAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 ttattgattccttgacttctttagctaatgaacatggtggagatcacgatgccgggcaag 1560
```

```
Qy    1561 AAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAATTG 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 aaattgaattagctctgaagataggaaaagtcaatgagtatgaaactgacgtcacaattg 1620

Qy    1621 ATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTG 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 ataaaggagggccaaagattttaattcttggagctggaagagtctgtcggccagctgctg 1680

Qy    1681 AGTTTCTGGCATCTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATC 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 agtttctggcatcttacccagacatatgtacctatggtgttgatgaccatgatgcagatc 1740

Qy    1741 AAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTA 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 aaattcatgttatcgtggcatctttgtatcaaaaagatgcagaagagacagttgatggta 1800

Qy    1801 TTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATCTTG 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 ttgaaaatacaactgctacccagcttgatgttgctgatattggaagcctttcagatcttg 1860

Qy    1861 TTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAG 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 tttctcaggttgaggttgtaattagcttgctgcctgctagttttcatgctgccattgcag 1920

Qy    1921 GAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCATGT 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 gagtatgcatagagttgaagaagcacatggtaacggcaagctatgttgatgaatccatgt 1980

Qy    1981 CAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCCTAG 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 caaacttgagccaagctgccaaagatgcaggtgtaactatactttgtgaaatgggcctag 2040

Qy    2041 ATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAAAGG 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 atcctggcatagatcacttgatgtcaatgaagatgattgatgaagctcatgcacgaaagg 2100

Qy    2101 GAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAAACA 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 gaaaaataaaggcatttacatcttactgtggtggattgccatctccagctgcagcaaaca 2160

Qy    2161 ATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAAATC 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 atccgcttgcctataaattcagttggaacccagctggtgcactccggtcagggaaaaatc 2220

Qy    2221 CTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATGAAT 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 ctgcagtctacaaatttcttggtgagacgatccatgtagatggtcataacttgtatgaat 2280

Qy    2281 CAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAAATC 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 cagcaaagaggctcagactacgagagcttccagcttttgctctggaacacttgccaaatc 2340

Qy    2341 GGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATATATA 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 ggaattccttgatatatggtgacctttatggtatctccaaagaagcatccaccatatata 2400

Qy    2401 GGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTGGGT 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 gggctactyttcgttacgaaggttttagtgagattatggtaaccctttccaaaactgggt 2460

Qy    2461 TCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATAAGGGTT 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 tctttgatgctgcaaatcatccactgctgcaagatactagtcgtccaacatataagggtt 2520

Qy    2521 TCCTTGATGAACTACTGAATAATATCTCCACAATTAACACGGACTTAGATATTGAAGCTT 2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 tccttgatgaactactgaataatatctccacaattaacacggacttagatattgaagctt 2580

Qy    2581 CTGGTGGATACGATGATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTTGCAAAAATA 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 ctggtggatacgatgatgacctgattgccagactgttgaagctcgggtgttgcaaaaata 2640

Qy    2641 AGGAAATAGCTGTTAAGACAGTCAAAACCATCAAGTTCTTGGGACTACATGAAGAGACTC 2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 aggaaatagctgttaagacagtcaaaaccatcaagttcttgggactacatgaagagactc 2700

Qy    2701 AAATACCTAAGGGTTGTTCGAGCCCATTTGATGTGATTTGCCAGCGAATGGAACAGAGGA 2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 aaatacctaagggttgttcgagcccatttgatgtgatttgccagcgaatggaacagagga 2760

Qy    2761 TGGCCTATGGCCACAATGAGCAAGACATGGTACTGCTCCACCACGAAGTCGAGGTGGAAT 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 tggcctatggccacaatgagcaagacatggtactgctccaccacgaagtcgaggtggaat 2820

Qy    2821 ACCCGGACGGGCAACCCGCCGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGAAGGTTG 2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 acccggacgggcaacccgccgaaaagcaccaagcgacgctactggagttcgggaaggttg 2880

Qy    2881 AAAATGGCAGGTCCACCACTGCCATGGCGCTGACCGTCGGCATTCCAGCAGCAATAGGGG 2940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2881 aaaatggcaggtccaccactgccatggcgctgaccgtcggcattccagcagcaatagggg 2940

Qy    2941 CCCTGCTATTGCTAAAGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGCAACCGG 3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2941 ccctgctattgctaaagaataaggtccagacgaaaggagtgatcaggcctctgcaaccgg 3000

Qy    3001 AAATCTACGTTCCAGCATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGAAAG 3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3001 aaatctacgttccagcattggagatcttggagtcgtcgggcatcaagctggttgagaaag 3060

Qy    3061 TGGAGACTTGAAAGTTCCCTGATACACAGATAAAGATAGTATGATATAGCAGGGCACATG 3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3061 tggagacttgaaagttccctgatacacagataaagatagtatgatatagcagggcacatg 3120

Qy    3121 TATCTTTTGTATTAACTCCGTTCTGGAATATATATTTGTGAACTAAAATGTGACAAATAA 3180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3121 tatcttttgtattaactccgttctggaatatatatttgtgaactaaaatgtgacaaataa 3180

Qy    3181 AAAGAACGGGTGGAGTATATTGTAAGAGACGGCAAAGAAACCTCTGTATATATGACCTGT 3240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3181 aaagaacgggtggagtatattgtaagagacggcaaagaaacctctgtatatatgacctgt 3240

Qy    3241 CGATATCAAATAATGCCGATCAGTT 3265
           |||||||||||||||||||||||||
Db    3241 cgatatcaaataatgccgatcagtt 3265


RESULT   2
AAV99564
ID   AAV99564 standard; cDNA to mRNA; 1908 BP.
XX
AC   AAV99564;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Maize lysine ketoglutarate reductase cDNA.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; soybean; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..1908
FT                   /*tag=  a
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
```

```
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   P-PSDB; AAW87762.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 3; Page 192-193; 231pp; English.
XX
CC   This is the nucleotide sequence of a truncated cDNA encoding a
CC   maize lysine ketoglutarate reductase (LKR) or saccharopine
CC   dehydrogenase (SDH) polypeptide (see AAW87762).  Isolated nucleic
CC   acid fragments comprising sequences encoding all or part of plant
CC   LKRs are new.  Also claimed are: (1) a chimeric gene comprising the
CC   fragment (or a subfragment) operably linked to a seed specific
CC   regulator, where the chimeric gene reduces LKR activity in plant
CC   seeds transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence 1908 BP; 537 A; 376 C; 439 G; 556 T; 0 other;

  Query Match             58.4%;  Score 1908;  DB 20;  Length 1908;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1908;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTGTGCCCGCCTTCTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 attgtgcccgccttctgctaggaggaggcaagaacggacctcgagtaaaccggattattg 60

Qy         61 TGCAGCCAAGCACAAGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 tgcagccaagcacaaggaggatccatcatgacgctcagtatgaggatgcaggatgcgaga 120

Qy        121 TTTCAGAAGACCTGTCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 tttcagaagacctgtcagaatgcggccttatcataggcatcaaacaacccaagctgcaga 180

Qy        181 TGATTCTTTCAGATAGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 tgattctttcagatagagcgtacgctttcttttcacacacacacaaagcccaaaaagaga 240

Qy        241 ATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 atatgccactgttagacaagatccttgaagaaagggtgtccttgtttgattatgagctaa 300

Qy        301 TTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ttgttggagatgatgggaaaagatcactagcatttgggaaatttgctggtagagctggac 360

Qy        361 TGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 tgatagatttcttacatggtctcggacagcgatatttgagccttggatactcgactccat 420

Qy        421 TTCTCTCTCTGGGACAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ttctctctctgggacaatctcatatgtatccttcgctcgctgcagccaaggctgcagtca 480

Qy        481 TTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATAGTGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ttgtcgttgcagaagagatagcaacatttggacttccatccggaatttgtccgatagtgt 540

Qy        541 TTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ttgtgttcactggagttggaaacgtctctcagggtgcgcaggagatattcaagttattgc 600

Qy        601 CCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 cccatacctttgttgatgctgagaagcttcccgaaatttttcaggccaggaatctgtcta 660

Qy        661 AGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 agcaatctcagtcgaccaagagagtatttcaactttatggttgtgttgtgacctctagag 720

Qy        721 ACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 acatagtttctcacaaggatcccaccagacaatttgacaaaggtgactattatgctcatc 780

Qy        781 CAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 cagaacactacacccctgtttttcatgaaagaattgctccatatgcatctgtcatcgtaa 840

Qy        841 ACTGTATGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 actgtatgtattgggagaagaggtttccaccattactaaatatggatcagttacagcaat 900

Qy        901 TGATGGAGACTGGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 tgatggagactggttgtcctttagtcggcgtttgtgacataacttgtgatattggaggtt 960

Qy        961 CCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ccattgaatttatcaacaagagtacatcaatagagaggcctttctttcggtatgatcctt 1020

Qy       1021 CTAAGAATTCATACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ctaagaattcataccatgatgatatggaaggtgccggagtggtctgcttggctgttgaca 1080

Qy       1081 TTCTCCCTACAGAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ttctccctacagaattctctaaagaggcctcccaacattttggaaacatactatctagac 1140

Qy       1141 TTGTTGCTAGTTTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ttgttgctagtttggcctcagtgaagcaaccggcagaacttccttcctacttgagaagag 1200

Qy       1201 CTTGCATTGCACATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 cttgcattgcacatgctggcagattaactcctttgtatgaatatatccctaggatgagaa 1260

Qy       1261 ATACTATGATAGATTTGGCACCCGCAAAAACAAATCCATTGCCTGACAAGAAGTATAGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 atactatgatagatttggcacccgcaaaaacaaatccattgcctgacaagaagtatagca 1320

Qy       1321 CCCTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ccctggtatctctcagtgggcacctatttgataagttccttataaatgaagctttggaca 1380

Qy       1381 TCATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 tcattgagacagctggaggttcatttcacttggttagatgtgaagttggacaaagcacgg 1440

Qy       1441 ATGATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 atgatatgtcatactcagagcttgaagtaggagcagatgatactgccacattggataaaa 1500

Qy       1501 TTATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCAAG 1560
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 ttattgattccttgacttctttagctaatgaacatggtggagatcacgatgccgggcaag 1560

Qy     1561 AAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAATTG 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 aaattgaattagctctgaagataggaaaagtcaatgagtatgaaactgacgtcacaattg 1620

Qy     1621 ATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 ataaaggagggccaaagattttaattcttggagctggaagagtctgtcggccagctgctg 1680

Qy     1681 AGTTTCTGGCATCTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATC 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 agtttctggcatcttacccagacatatgtacctatggtgttgatgaccatgatgcagatc 1740

Qy     1741 AAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTA 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 aaattcatgttatcgtggcatctttgtatcaaaaagatgcagaagagacagttgatggta 1800

Qy     1801 TTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATCTTG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 ttgaaaatacaactgctacccagcttgatgttgctgatattggaagcctttcagatcttg 1860

Qy     1861 TTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATG 1908
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 tttctcaggttgaggttgtaattagcttgctgcctgctagttttcatg 1908


RESULT   3
AAC42953
ID   AAC42953 standard; DNA; 3195 BP.
XX
AC   AAC42953;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 37480.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
```

```
PR   23-JUL-1999;    99US-0145224.
PR   26-JUL-1999;    99US-0145276.
PR   27-JUL-1999;    99US-0145913.
PR   27-JUL-1999;    99US-0145918.
PR   27-JUL-1999;    99US-0145919.
PR   28-JUL-1999;    99US-0145951.
PR   02-AUG-1999;    99US-0146386.
PR   02-AUG-1999;    99US-0146388.
PR   02-AUG-1999;    99US-0146389.
PR   03-AUG-1999;    99US-0147038.
PR   04-AUG-1999;    99US-0147204.
PR   04-AUG-1999;    99US-0147302.
PR   05-AUG-1999;    99US-0147192.
PR   05-AUG-1999;    99US-0147260.
PR   06-AUG-1999;    99US-0147303.
PR   06-AUG-1999;    99US-0147416.
PR   09-AUG-1999;    99US-0147493.
PR   09-AUG-1999;    99US-0147935.
PR   10-AUG-1999;    99US-0148171.
PR   11-AUG-1999;    99US-0148319.
PR   12-AUG-1999;    99US-0148341.
PR   13-AUG-1999;    99US-0148565.
PR   13-AUG-1999;    99US-0148684.
PR   16-AUG-1999;    99US-0149368.
PR   17-AUG-1999;    99US-0149175.
PR   18-AUG-1999;    99US-0149426.
PR   20-AUG-1999;    99US-0149722.
PR   20-AUG-1999;    99US-0149723.
PR   20-AUG-1999;    99US-0149929.
PR   23-AUG-1999;    99US-0149902.
PR   23-AUG-1999;    99US-0149930.
PR   25-AUG-1999;    99US-0150566.
PR   26-AUG-1999;    99US-0150884.
PR   27-AUG-1999;    99US-0151065.
PR   27-AUG-1999;    99US-0151066.
PR   27-AUG-1999;    99US-0151080.
PR   30-AUG-1999;    99US-0151303.
PR   31-AUG-1999;    99US-0151438.
PR   01-SEP-1999;    99US-0151930.
PR   07-SEP-1999;    99US-0152363.
PR   10-SEP-1999;    99US-0153070.
PR   13-SEP-1999;    99US-0153758.
PR   15-SEP-1999;    99US-0154018.
PR   16-SEP-1999;    99US-0154039.
PR   20-SEP-1999;    99US-0154779.
PR   22-SEP-1999;    99US-0155139.
PR   23-SEP-1999;    99US-0155486.
PR   24-SEP-1999;    99US-0155659.
PR   28-SEP-1999;    99US-0156458.
PR   29-SEP-1999;    99US-0156596.
PR   04-OCT-1999;    99US-0157117.
PR   05-OCT-1999;    99US-0157753.
PR   06-OCT-1999;    99US-0157865.
PR   07-OCT-1999;    99US-0158029.
PR   08-OCT-1999;    99US-0158232.
PR   12-OCT-1999;    99US-0158369.
PR   13-OCT-1999;    99US-0159293.
PR   13-OCT-1999;    99US-0159294.
PR   13-OCT-1999;    99US-0159295.
PR   14-OCT-1999;    99US-0159329.
PR   14-OCT-1999;    99US-0159330.
PR   14-OCT-1999;    99US-0159331.
PR   14-OCT-1999;    99US-0159637.
PR   14-OCT-1999;    99US-0159638.
PR   18-OCT-1999;    99US-0159584.
PR   21-OCT-1999;    99US-0160741.
PR   21-OCT-1999;    99US-0160767.
PR   21-OCT-1999;    99US-0160768.
PR   21-OCT-1999;    99US-0160770.
PR   21-OCT-1999;    99US-0160814.
PR   21-OCT-1999;    99US-0160815.
PR   22-OCT-1999;    99US-0160980.
PR   22-OCT-1999;    99US-0160981.
PR   22-OCT-1999;    99US-0160989.
PR   25-OCT-1999;    99US-0161404.
PR   25-OCT-1999;    99US-0161405.
PR   25-OCT-1999;    99US-0161406.
PR   26-OCT-1999;    99US-0161359.
PR   26-OCT-1999;    99US-0161360.
PR   26-OCT-1999;    99US-0161361.
PR   28-OCT-1999;    99US-0161920.
PR   28-OCT-1999;    99US-0161992.
PR   28-OCT-1999;    99US-0161993.
PR   29-OCT-1999;    99US-0162142.

  Query Match             37.8%;  Score 1233.2;  DB 21;  Length 3195;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 1997;  Conservative    1;  Mismatches 1029;  Indels   67;  Gaps    8;

Qy     15 CTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTGTGCAGCCAAGCACA 74
          ||  ||   || || || || || |  |  |   ||| ||| | || ||||||    |
Db    121 cttttacacggtgggaaagacagaaccggcatttcccgcattgtggttcagccatctgct 180

Qy     75 AGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGATTTCAGAAGACCTG 134
          | | | ||||||||||| ||   |||||| ||||  || || || ||||| || ||  ||
Db    181 aagcgtatccatcatgatgccttgtatgaagatgttgggtgtgaaatttctgatgatttg 240

Qy    135 TCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGATGATTCTTTCAGAT 194
          || || || || |||||  | || |||||||||||  ||||  | ||||||||| ||||
Db    241 tctgattgtgggcttatacttggaatcaaacaacctgagctagaaatgattcttccagag 300

Qy    195 AGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGAATATGCCACTGTTA 254
          ||||| ||||||||||||||||| || || || || || |||||||| |||||  ||||
Db    301 agagcatacgctttcttttcacatactcataaggcacagaaagagaacatgcctttgttg 360

Qy    255 GACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAATTGTTGGAGATGAT 314
          || || || |||   || || ||| | |||| |||||||||||| |||||||| ||| ||
Db    361 gataaaattctttctgagagagtgactttgtgtgattatgagctcattgttggggatcat 420

Qy    315 GGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGACTGATAGATTTCTTA 374
          |||||| ||| | | || ||||| |||| ||| || |||||||| ||  | || ||||||
Db    421 gggaaacgattattggcgtttggtaaatatgcaggcagagctggtcttgttgacttctta 480

Qy    375 CATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCATTTCTCTCTCTGGGA 434
          || || || ||||||||  |  | |  || |||||||| || || || ||||| || ||
Db    481 cacggacttggacagcgtaagctcattctaggatactcaacacctttcctctcgctcggt 540

Qy    435 CAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCATTGTCGTTGCAGAA 494
            |||  |||||||| | ||  | ||||| || || || || || |||   ||||  |||
Db    541 gcatcgtatatgtattcctcattggctgctgcaaaagccgctgtaatttctgttggtgaa 600

Qy    495 GAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATAGTGTTTGTGTTCACTGGA 554
          || || ||||     ||||| ||||  ||||| || ||  | || ||||| ||||| |||
Db    601 gaaattgcaagccagggactgccattaggaatctgccctcttgtatttgtcttcaccgga 660

Qy    555 GTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGCCCCATACCTTTGTT 614
             ||||| || |||| ||| ||||| || || |||||| |  | || || || ||||||
Db    661 acaggaaatgtttctctgggggcgcaagaaattttcaagcttcttcctcacactttgtt 720

Qy    615 GATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTAAGCAATCTCAG--- 671
          ||  |    || ||||| ||| | |||         ||      ||  |||  |  |
Db    721 gaaccaagcaaacttcctgaactatttgtaaaagacaaaggaattagtcaaaatgggatt 780

Qy    672 TCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAGACATAGTTTCT 731
          || || ||| |||| | |||| | ||||||||| || | |||     |||||| |||
Db    781 tcaacaaagcgagtctatcaagtatatggttgtattattaccagccaagacatggttgaa 840

Qy    732 CACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATCCAGAACACTAC 791
          ||||| ||||| | |    ||| |||||||  ||||||||||| || || ||||| |||
Db    841 cacaaagatccatcaaagtcattcgacaaagccgactattatgcacacccggaacattac 900

Qy    792 ACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAAACTGTATGTAT 851
          |  || ||||| || ||||  ||  | |||||| | |||||  | ||||||||||||||
Db    901 aatccagttttccacgaaaagatatcgccatatacgtctgttcttgtaaactgtatgtac 960

Qy    852 TGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAATTGATGGAGACT 911
```

```
Db     961 tgggagaagaggtttccctgtcttctgagcacaaaacagcttcaagatttaacaaaaaaa 1020
Qy     912 GGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTTCCATTGAATTT 971
Db    1021 ggactcccactagtaggcatatgtgatataacttgtgacatcggtggctccattgaattt 1080
Qy     972 ATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTTCTAAGAATTCA 1031
Db    1081 gttaaccgagctactttaatcgattcccctttcttcaggtttaatccctcgaacaattca 1140
Qy    1032 TACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACATTCTCCCTACA 1091
Db    1141 tactacgatgacatggatggggatggcgtactatgcatggctgttgacattttacccaca 1200
Qy    1092 GAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGACTTGTTGCTAGT 1151
Db    1201 gaatttgcaaaagaggcatcccagcattttggagatattctttccggatttgtcggtagt 1260
Qy    1152 TTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAGCTTGCATTGCA 1211
Db    1261 ttggcttcaatgactgaaatttcagatctaccagcacatctgaagagggcttgcataagc 1320
Qy    1212 CATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAAATACTATGATA 1271
Db    1321 tataggggagaattgacatctttgtatgagtatattccacgtatgaggaagtcaaatcca 1380
Qy    1272 GATTTGGCACCCGCAAAAACAAATCCATTGCCTG---------ACAAGAAGTATAGCACC 1322
Db    1381 gaagaggcacaagataatattatcgccaacggggtttccagccagagaacattcaacata 1440
Qy    1323 CTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACATC 1382
Db    1441 ttggtatctctgagcggacacctatttgataagtttctgataaacgaagctcttgatatg 1500
Qy    1383 ATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGGAT 1442
Db    1501 atcgaagcggctggtggctcatttcatttggctaaatgtgaactggggcagagcgctgat 1560
Qy    1443 GATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAATT 1502
Db    1561 gctgaatcgtactcagaacttgaagttggtgcggatgataagagagtattggatcaaatc 1620
Qy    1503 ATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGC---CGGGCAA 1559
Db    1621 attgattcattaactcggttagctaatccaaatgaagattatatatccccacatagagaa 1680
Qy    1560 GAAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAG--------TATGAAACTGACG 1611
Db    1681 gcaaataagatctcactgaagattggtaaagtccagcaagaaaatgagataaaagagaag 1740
Qy    1612 TCACAATTGATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGC 1671
Db    1741 cctgaaatgacgaa--aaaatcaggtgttttgattcttggtgctggacgtgtgtgtcgcc 1798
Qy    1672 CAGCTGCTGAGTTTCTGGCATCTTACCCAGACATATGTAC-------------------- 1711
Db    1799 cagctgctgatttcctagcttcagttagaaccatttcgtcacagcaatggtacaaaacat 1858
Qy    1712 -CTATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATCTTTGTATC 1770
Db    1859 atttcggagcagactctgaagagaaaacagatgttcatgtgattgtcgcgtctctgtatc 1918
Qy    1771 AAAAAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCAGCTTGATG 1830
Db    1919 ttaaggatgccaaagagacggttgaaggtatttcagatgtagaagcagttcggctagatg 1978
Qy    1831 TTGCTGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGC 1890
Db    1979 tatctgatagtgaaagtctccttaagtatgtttctcaggttgatgttgtcctaagtttat 2038
Qy    1891 TGCCTGCTAGTTTTCATGCTGCCATTGCAGGAGTATGCATAGAGTTGAAGAAGCACATGG 1950
```

```
Db    2039 tacctgcaagttgtcatgctgttgtagcaaagacatgcattgagctgaagaagcatctcg 2098
Qy    1951 TAACGGCAAGCTATGTTGATGAATCCATGTCAAACTTGAGCCAAGCTGCCAAAGATGCAG 2010
Db    2099 tcactgctagctatgttgatgatgaaacgtccatgttacatgagaaggctaagagtgctg 2158
Qy    2011 GTGTAACTATACTTTGTGAAATGGGCCTAGATCCTGGCATAGATCACTTGATGTCAATGA 2070
Db    2159 ggataacgattctaggcgaaatgggactggaccctggaatcgatcacatgatggcgatga 2218
Qy    2071 AGATGATTGATGAAGCTCATGCACGAAAGGGAAAAATAAAGGCATTTACATCTTACTGTG 2130
Db    2219 aaatgatcaacgatgctcatatcaaaaaagggaaagtgaagtcttttacctcttattgtg 2278
Qy    2131 GTGGATTGCCATCTCCAGCTGCAGCAAACAATCCGCTTGCCTATAAATTCAGTTGGAACC 2190
Db    2279 gagggcttccctctcctgctgcagcaaataatccattagcatataaatttagctggaacc 2338
Qy    2191 CAGCTGGTGCACTCCGGTCAGGGAAAAATCCTGCAGTCTACAAATTTCTTGGTGAGACGA 2250
Db    2339 ctgctggagcaattcgagctggtcaaaaccccgccaaatacaaaagcaacggcgacataa 2398
Qy    2251 TCCATGTAGATGGTCATAACTTGTATGAATCAGCAAAGAGGCTCAGACTACGAGAGCTTC 2310
Db    2399 tacatgttgatgggaagaatctctatgattccgcggcaagattccgagtacctaatcttc 2458
Qy    2311 CAGCTTTTGCTCTGGAACACTTGCCAAATCGGAATTCCTTGATATATGGTGACCTTTATG 2370
Db    2459 cagcttttgcattggagtgttttccaaatcgtgactccttggtttacggggaacattatg 2518
Qy    2371 GTATCTCCAAAGAAGCATCCACCATATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTG 2430
Db    2519 gcatcgagagcgaagcaacaacgatatttcgtggaacactcagatatgaagggtttagta 2578
Qy    2431 AGATTATGGTAACCCTTTCCAAAACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGC 2490
Db    2579 tgataatggcaacactttcgaaacttggattctttgacagtgaagcaaatcaagtactct 2638
Qy    2491 AAGATACTAGTCGTCCAACATATAAGGGTTTCCTTGATGAACTACTGAATAATATCTCCA 2550
Db    2639 ccactggaaagaggattacgtttggtgctcttttaagtaacattctaaataaggatgcag 2698
Qy    2551 CAATTAACACGGACTTAGATATTGAAGCTTCTGGTGGATACGATGATGACCTGATTGCCA 2610
Db    2699 acaatgaatc---------------agagcccctagcgggagaagaagagataagcaaga 2743
Qy    2611 GACTGTTGAAGCTCGGGTGTTGCAAAAATAAGGAAATAGCTGTTAAGACAGTCAAAACCA 2670
Db    2744 gaattatcaagcttggacattcc------aaggagactgcagccaaagctgccaaaacaa 2797
Qy    2671 TCAAGTTCTTGGGACTACATGAAGAGACTCAAATACCTAAGGGTTGTTCGAGCCCATTTG 2730
Db    2798 ttgtattcttggggttcaacgaagagagggaggttccatcactgtgtaaaagcgtatttg 2857
Qy    2731 ATGTGATTTGCCAGCGAATGGAACAGAGGATGGCCTATGGCCACAATGAGCAAGACATGG 2790
Db    2858 atgcaacttgttacctaatggaagagaaactagcttattccggaaatgaacaggacatgg 2917
Qy    2791 TACTGCTCCACCACGAAGTCGAGGTGGAATACCCGGACGGGCAACCCGCCGAAAAGCACC 2850
Db    2918 tgcttttgcatcacgaagtagaagtggaattccttgaaagcaaacgtatagagaagcaca 2977
Qy    2851 AAGCGACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCACCACTGCCATGGCGC 2910
Db    2978 ctgcgactcttttggaattcggggacatcaagaatggacaaacaacaaccgctatggcca 3037
Qy    2911 TGACCGTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAGGTCCAGA 2970
Db    3038 agactgttgggatccctgcagccattggagctctgctgttaattgaagacaagatcaaga 3097
Qy    2971 CGAAAGGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAGCATTGGAGATCTTGG 3030
Db    3098 caagaggagtcttaaggcctctcgaagcagaggtgtatttgccagctttggatatattgc 3157
```

```
Qy   3031 AGTCGTCGGGCATCAAGCTGGTTGAGAAAGTGGA 3064
          |  | |  || || |||||| | ||||| |  ||
Db   3158 aagcatatggtataaagctgatggagaaggcaga 3191

RESULT   4
AAV99562
ID   AAV99562 standard; cDNA to mRNA; 2582 BP.
XX
AC   AAV99562;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Soybean lysine ketoglutarate reductase cDNA.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; soybean; ss.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..2357
FT                   /*tag=  a
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;    98WO-US06051.
XX
PR   27-MAR-1997;    97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   P-PSDB; AAW87760.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 3; Page 182-184; 231pp; English.
XX
CC   This is the nucleotide sequence of a near full-length cDNA encoding
CC   soybean lysine ketoglutarate reductase (LKR, see AAW87760), or
CC   saccharopine dehydrogenase (SDH).  It was obtained by PCR
CC   amplification of developing seed cDNA using primers (see AAV99560-61)
CC   based on Arabidopsis LKR/SDH sequences, and by RACE.  Isolated
CC   nucleic acids comprising sequences encoding all or part of plant
CC   LKRs are new.  Also claimed are: (1) a chimeric gene comprising the
CC   fragment (or a subfragment) operably linked to a seed specific
CC   regulator, where the chimeric gene reduces LKR activity in plant
CC   seeds transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence 2582 BP; 798 A; 482 C; 562 G; 739 T; 1 other;

  Query Match             27.8%;  Score 907;  DB 20;  Length 2582;
  Best Local Similarity   62.7%;  Pred. No. 4.2e-247;
  Matches 1531;  Conservative    2;  Mismatches  857;  Indels   51;  Gaps    6;

Qy    727 TTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATCCAGAAC 786
          ||   | ||| |||| |   | |   ||||||||||  ||||| |||   || || || |
Db      1 ttgaacccaaagatcacgtgatagtgtttgacaaagcagactactattcacaccctgagc 60

Qy    787 ACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAAACTGTA 846
          | ||||  ||   ||| ||||||| ||| || ||||||||||||||| || || || || |
Db     61 attacaatcccactttccatgaaaaaatagcaccatatgcatctgttattgtcaattgca 120

Qy    847 TGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAATTGATGG 906
          ||||||||||||| || ||||| | ||| |  |      | ||| | ||  | || ||||
Db    121 tgtattgggagaaaagatttcctcaattgccgagctataagcagatgcaagacttaatgg 180

Qy    907 AGACTGGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTTCCATTG 966
               ||  | ||  | || ||  |   ||||||||| |||||| || ||||| ||||
Db    181 gccgggggagcccccttgttggaatagctgacataacgtgtgatatagggggttcaattg 240

Qy    967 AATTTATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTTCTAAGA 1026
          | ||| | |||    |||| ||||| ||    || |||||  | |||||||| |  |  |
Db    241 agtttgttaaccgcggtacttcaattgattcacccttcttcagatatgatcccttaacaa 300

Qy   1027 ATTCATACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACATTCTCC 1086
          |||| ||||||||||||||||| ||    |||||| | ||||| |||||||||||||| |
Db    301 attcctaccatgatgatatggaggggaatggagtgatatgcttagctgttgacattcttc 360

Qy   1087 CTACAGAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGACTTGTTG 1146
          | ||||||||  | || ||||| |||||||||||||||||||||| ||    | ||||||
Db    361 caacagaatttgcaaaggaggcttcccaacattttggaaacatactttcccaatttgttg 420

Qy   1147 CTAGTTTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAGCTTGCA 1206
            | |||||| || |  |   |     || |  | ||| |  |||| || ||||||||||
Db    421 taaatttggcttctgctacagacattacaaagttgcctgctcacttaaggagagcttgca 480

Qy   1207 TTGCACATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAA------ 1260
          | || |||   ||     |||| |  || ||||| ||||||||  | ||| | |
Db    481 tagcccataaaggagtgctaacctccttatatgattatatcccacgcatgcggagttctg 540

Qy   1261 ATACTATGATAGATTTGGCACCCGCAAAAACAAATCCATTGCCTGACAAGAAGTATAGCA 1320
          || |   |  ||  |  | |  |||| |||    || ||      | | |||||| |  |
Db    541 attcagaggaagtatcagaaaacgcagaaaattctctatccaacaaaaggaagtacaata 600

Qy   1321 CCCTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACA 1380
              ||| ||||| ||||| ||| |||||||| |||| ||  ||||||||  || || || |
Db    601 tatcggtgtctctgagtggtcacttatttgatcagtttctgataaatgaggccttagata 660

Qy   1381 TCATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGG 1440
          | |||||  | || ||||| || || |||||| || |  ||  | || || || ||||  |
Db    661 ttattgaagctgcaggaggctccttccacttagtcaactgccatgtgggtcagagcattg 720

Qy   1441 ATGATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAA 1500
          | |   | |||| ||| || |||||||| || |||||  | |  ||      ||||| |||
Db    721 aagccgtatcattctctgaacttgaagttggtgcagataacagggctgttctggatcaaa 780

Qy   1501 TTATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCAAG 1560
          | |||||||| || ||  || | |||| | |  ||      |||     | |    ||||
Db    781 tcattgattctttaactgctattgctagtccaactgaacatgatagattttcaaatcaag 840

Qy   1561 A---AATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAA 1617
          |   || | ||| |  | || ||| | || |||||  | ||| |||  |  ||      |
Db    841 attcaagtaaaatttcacttaagcttggtaaagttgaagagaatggcatagagaaggaat 900

Qy   1618 TTGA---TAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAG 1674
           |||    | |     | |   | ||||||||||||||||||||||  | |||||||  ||||
Db    901 ctgaccccagaaagaaggctgcggttttaattcttggagctggtcgggtctgtcaaccag 960

Qy   1675 CTGCTGAGTTTCTGGCATCTTACCCAGACATATGTAC--------------------CT 1713
```

```
            |||||||  |  |  |||| |    |     ||   |                      |
Db      961 ctgctgaaatgttatcatcatttggaaggccatcatcgagccaatggtataaaacattgt 1020

Qy     1714 ATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATCTTTGTATCAAA 1773
              |  | ||||    |   |  |  | |  |  | || || |||| |||| |||| |  |
Db     1021 tggaagatgattttgaatgtcaaactgatgtagaagtcattgtgggatctctgtacctga 1080

Qy     1774 AAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCAGCTTGATGTTG 1833
            | ||||||||  |||| ||||| || |||  ||||  ||| |  |  |||||||||||
Db     1081 aggatgcagagcagactgttgagggcattccaaatgtaaccggaattcagcttgatgtga 1140

Qy     1834 CTGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGC 1893
              |||  ||  |   | |   |     |||| |||||||| ||||| || || |||||||
Db     1141 tggatcgtgccaatttgtgtaagtacatttcacaggttgacgttgttataagtttgctgc 1200

Qy     1894 CTGCTAGTTTTCATGCTGCCATTGCAGGAGTATGCATAGAGTTGAAGAAGCACATGGTAA 1953
            |  | |||| ||||  |    | |||   |  ||||| ||| |||| || ||  | || |
Db     1201 ccccaagttgtcatattattgtagcaaatgcttgcattgagctgaaaaacatcttgtca 1260

Qy     1954 CGGCAAGCTATGTTGATGAATCCATGTCAAACTTGAGCCAAGCTGCCAAAGATGCAGGTG 2013
            | || ||||||||||||   ||||||||||   | |   |    || |||||||| ||
Db     1261 ctgctagctatgttgatagctccatgtcaatgctaaatgataaggctaaagatgctggca 1320

Qy     2014 TAACTATACTTTGTGAAATGGGCCTAGATCCTGGCATAGATCACTTGATGTCAATGAAGA 2073
            |||| || ||| | || |||||| | || || || || | |||  ||||| |||||||||
Db     1321 taacaattcttggagagatgggcttggacccaggaattggtcatatgatggcaatgaaga 1380

Qy     2074 TGATTGATGAAGCTCATGCACGAAAGGGAAAAATAAAGGCATTTACATCTTACTGTGGTG 2133
            ||||  |  |||| ||||   | ||||| ||||||||| | || || ||||| |||||||
Db     1381 tgatcaaccaagcacatgtgaggaaggggaaaataaagtctttcacttcttattgtggtg 1440

Qy     2134 GATTGCCATCTCCAGCTGCAGCAAACAATCCGCTTGCCTATAAATTCAGTTGGAACCCAG 2193
            || | |||||||| |  || || ||||||||  | || ||||||||||||||||| || |
Db     1441 gacttccatctcctgaagctgctaacaatccattagcatataaattcagttggaatcctg 1500

Qy     2194 CTGGTGCACTCCGGTCAGGGAAAAATCCTGCAGTCTACAAATTTCTTGGTGAGACGATCC 2253
            | || ||  ||||  | |||   ||||||||   ||||||||    |||||| ||  | |
Db     1501 caggagccatccgagctgggcgcaatcctgccacctacaaatggggtggtgaaactgtac 1560

Qy     2254 ATGTAGATGGTCATAACTTGTATGAATCAGCAAAGAGGCTCAGACTACGAGAGCTTCCAG 2313
            || | |||||  |  |  | ||||| || || |  || || || ||||  || ||||| |
Db     1561 atattgatggggacgatctttatgattcggctacaagactaaggctaccggaccttcctg 1620

Qy     2314 CTTTTGCTCTGGAACACTTGCCAAATCGGAATTCCTTGATATATGGTGACCTTTATGGTA 2373
            |||||||| |||||    | |||||||| ||||| ||  | ||||| ||  | ||||| |
Db     1621 ctttttgctttggaatgtctcccaaatcgcaattcattactttatggggatttgtatggaa 1680

Qy     2374 TCTCCAAAGAAGCATCCACCATATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTGAGA 2433
            |    |  |||||||| ||||| |   | |  || :| || || ||||| ||||||||||
Db     1681 t---aactgaagcatcaaccatttccgtggaaccctccgctatgaaggatttagtgaga 1737

Qy     2434 TTATGGTAACCCTTTCCAAAACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGCAAG 2493
            | ||||  || || || |  | | | || ||| |   || |   ||| |  ||||
Db     1738 tcatggggacactgtctaggattagcttatttaacaatgaagcccattcgttgctaatga 1797

Qy     2494 ATACTAGTCGTCCAACATATAAGGGTTTCCTTGATGAACTACTGAATAATATCTCCACAA 2553
            ||       | ||||| |  ||    ||| |   |||||| || ||   | |       |
Db     1798 atggacaaagaccaactttcaaaaaattcttatttgaacttctcaaagttgttggtgata 1857

Qy     2554 TTAACACGGACTTAGATATTGAAGCTTCTGGTGGATACGATGATGACCTGATTGCCAGAC 2613
             |      ||  ||   || | |               ||  ||||| | || |    |
Db     1858 atccagatgaactattgatagga---------------gagaatgacatcatggagcaaa 1902

Qy     2614 TGTTGAAGCTCGGGTGTTGCAAAAATAAGGAAATAGCTGTTAAGACAGTCAAAACCATCA 2673
            | || |  |  |||   |||||| || |   ||  ||  |  ||||||  ||||| ||||
Db     1903 tattaatacaagggcactgcaaagatcaaagaacggcaatggagacagcaaaaacaatca 1962

Qy     2674 AGTTCTTGGGACTACATGAAGAGACTCAAATACCTAAGGGTTGTTCGAGCCCATTTGATG 2733
              ||||||||||| | |||  | ||| |||| |||      ||    ||  | |||||||
Db     1963 ttttcttgggacttcttgaccaaactgaaatccctgcttcctgcaaaagtgcttttgatg 2022

Qy     2734 TGATTTGCCAGCGAATGGAACAGAGGATGGCCTATGGCCACAATGAGCAAGACATGGTAC 2793
            |   |||    || |||||  ||||| |  | ||   |  ||  ||  | || ||||| |
Db     2023 ttgcttgtttccgcatggaggagaggttatcatacaccagcacagaaaaggatatggtgc 2082

Qy     2794 TGCTCCACCACGAAGTCGAGGTGGAATACCCGGACGGGCAACCCGCCGAAAAGCACCAAG 2853
            |  | || || ||||| ||  | |||||||| ||  | |||    | || |||||   ||
Db     2083 ttttgcatcatgaagtggaaatagaatacccagatagccaaattacagagaagcatagag 2142

Qy     2854 CGACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCACCACTGCCATGGCGCTGA 2913
            | ||  |||| || || ||||||  |    |||  |   |||| ||||||||||| || |
Db     2143 ctactttacttgaatttgggaagactcttgatgaaaaaccacaactgccatggccctta 2202

Qy     2914 CCGTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAGGTCCAGACGA 2973
            | || || |||||||| ||  | || ||  |||| ||  | |  || ||  | ||||| |
Db     2203 ctgttggtattccagctgctgttggagctttgcttttattgacaaacaaaattcagacaa 2262

Qy     2974 AAGGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAGCATTGGAGATCTTGGAGT 3033
             ||||||  | |||||| |  |||| ||| | |||  ||||||| |||| ||  | ||
Db     2263 gaggagtcttaaggcctatcgaacctgaagtatacaatccagcactggatattatagaag 2322

Qy     3034 CGTCGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGAAAGTTCCCTGATACACAGATAA 3093
            | |  || |||||| || | |||||    |||   |     |:     ||  |   |||
Db     2323 cttatgggatcaagttgatagagaagaccgagtaatttgcatytatgaattgatgtatag 2382

Qy     3094 AGATAGTATGATATAGCAGGGCACATGTATCTTTTGTATTA 3134
               ||   | || ||         |||| |  |||| || |
Db     2383 gtgtacattaatgtacaccatgcaatgtttgatttgaataa 2423

RESULT   5
AAV99565
ID   AAV99565 standard; cDNA to mRNA; 720 BP.
XX
AC   AAV99565;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Rice lysine ketoglutarate reductase cDNA.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; rice; ss.
XX
OS   Oryza sativa.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..720
FT                   /*tag=  a
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   P-PSDB; AAW87763.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 3; Page 196; 231pp; English.
XX
CC   This is the nucleotide sequence of a partial cDNA encoding a
CC   rice lysine ketoglutarate reductase (LKR) or saccharopine
CC   dehydrogenase (SDH) polypeptide (see AAW87763).  It was isolated
CC   from an expressed sequence tag database by comparison to an
CC   Arabidopsis LKR/SDH protein (see AAW87759).  Isolated nucleic
```

```
CC   acid fragments comprising sequences encoding all or part of plant
CC   LKRs are new.  Also claimed are: (1) a chimeric gene comprising the
CC   fragment (or a subfragment) operably linked to a seed specific
CC   regulator, where the chimeric gene reduces LKR activity in plant
CC   seeds transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence 720 BP; 203 A; 150 C; 158 G; 206 T; 3 other;

  Query Match             16.0%;  Score 522.6;  DB 20;  Length 720;
  Best Local Similarity   85.5%;  Pred. No. 4.7e-138;
  Matches  579;  Conservative    1;  Mismatches   97;  Indels    0;  Gaps    0;

Qy       1848 CTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCAT 1907
              |||||  ||||||||||||||||||| || ||| |||||||||||||||| |||||||||
Db         11 ctttccaatcttgtttctcaggttgaagtagtagttagcttgctgcctgccagttttcat 70

Qy       1908 GCTGCCATTGCAGGAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTT 1967
              |||||||| ||| |||||||||||||| ||||||||||| |||| || ||||||||||||
Db         71 gctgccatagcaagagtatgcatagagatgaagaagcacttggtcactgcaagctatgtt 130

Qy       1968 GATGAATCCATGTCAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGT 2027
              ||||| ||||||||||| |||   ||| ||||  ||| ||| ||||||||||| || |||
Db        131 gatgagtccatgtcaaagttggaacaatctgcagaaggtgctggtgtaactattctctgt 190

Qy       2028 GAAATGGGCCTAGATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCT 2087
              ||||||||||| |||||||||||| ||||  |||||||||||||||||||||| ||||| 
Db        191 gaaatgggcctggatcctggcatanatcatatgatgtcaatgaagatgattgacgaagca 250

Qy       2088 CATGCACGAAAGGGAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCA 2147
              ||| |||| ||||| ||||||||| |||||||||| |  ||||| ||| | |||||||||
Db        251 cattcacggaaggggaaaataaagtcatttacatccttttgtggaggacttccatctcca 310

Qy       2148 GCTGCAGCAAACAATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGG 2207
              ||| | ||||||||||| ||||| ||||| ||||||||||  |||||||||||  ||||
Db        311 gcttctgcaaacaatccacttgcttataagttcagttggagtccagctggtgccatccgt 370

Qy       2208 TCAGGGAAAAATCCTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCAT 2267
               |||||| ||| ||||| ||||||||||||| ||| || |  ||||||||||||||| ||
Db        371 gcagggagaaaccctgctgtctacaaatttcatggagaaatcatccatgtagatggtgat 430

Qy       2268 AACTTGTATGAATCAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAA 2327
              || ||||||||||| ||||||||||||||| ||| ||| |||||||||||||| ||||||
Db        431 aaattgtatgaatccgcaaagaggctcagattacmagaacttccagcttttgcactggaa 490

Qy       2328 CACTTGCCAAATCGGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCA 2387
              ||||||||||| |||||||||||||| ||||| ||||| ||||| |||||||||||||||
Db        491 cacttgccaaaccggaattccttgatgtatggagacctgtatgggatctccaaagaagca 550

Qy       2388 TCCACCATATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTT 2447
              || ||  | || |||||||||:||||||| ||||| |||| |||||| |||| |||| |
Db        551 tctactgtgtacagggctactcttcgttatgaaggatttaatgagataatggcaaccttc 610

Qy       2448 TCCAAAACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCA 2507
               | |||| |||||| ||||||||||||| |||||||||| ||||| | |||| ||| ||
Db        611 gcgaaaattgggttttttgatgctgcaagtcatccactgttgcaacaaactactcgccct 670
```

```
Qy       2508 ACATATAAGGGTTTCCT 2524
              ||||| | || ||||||
Db        671 acatacanggatttcct 687

RESULT   6
AAV99568
ID   AAV99568 standard; DNA; 1449 BP.
XX
AC   AAV99568;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Arabidopsis saccharopine dehydrogenase coding region.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   P-PSDB; AAW87766.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 3; Page 200; 231pp; English.
XX
CC   This nucleotide sequence of a coding region for an Arabidopsis
CC   thaliana polypeptide (see AAW87766) comprising the saccharopine
CC   dehydrogenase (SDH) domain of the lysine ketoglutarate reductase/
CC   SDH protein.  Isolated nucleic acid fragments comprising sequences
CC   encoding all or part of plant LKR polypeptides are new.  Also
CC   claimed are: (1) a chimeric gene comprising the fragment (or a
CC   subfragment) operably linked to a suitable seed specific regulator,
CC   where the chimeric gene reduces LKR activity in plant seeds
CC   transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence 1449 BP; 446 A; 267 C; 356 G; 380 T; 0 other;

  Query Match             15.8%;  Score 515;  DB 20;  Length 1449;
  Best Local Similarity   61.7%;  Pred. No. 1e-135;
```

```
  Matches  893;  Conservative    1;  Mismatches  511;  Indels   42;  Gaps    3;

Qy   1639 TTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTGAGTTTCTGGCATCTTACC 1698
          |||| |||||||| |||||| | || ||||| ||||||||||| || || || ||
Db     20 ttttgattcttggtgctggacgtgtgtgtcgcccagctgctgatttcctagcttcagtta 79

Qy   1699 CAGACATATGTACCTAT---------------------GGTGTTGATGACCATGATGCAG 1737
           |  ||| |   |  |                      || |  ||     | ||   |
Db     80 gaaccatttcgtcacagcaatggtacaaaacatatttcggagcagactctgaagagaaaa 139

Qy   1738 ATCAAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATG 1797
             |  ||||||| || || || ||| |||||| || |||||  ||||||| ||||| |
Db    140 cagatgttcatgtgattgtcgcgtctctgtatcttaaggatgccaaagagacggttgaag 199

Qy   1798 GTATTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATC 1857
          |||||  | ||  |   ||    | ||| |||||  |||||| || ||| ||     |
Db    200 gtatttcagatgtagaagcagttcggctagatgtatctgatagtgaaagtctccttaagt 259

Qy   1858 TTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTG 1917
           ||||||||||||||| |||||  | || ||  | ||||| |||| ||||||||   | |
Db    260 atgtttctcaggttgatgttgtcctaagtttattacctgcaagttgtcatgctgttgtag 319

Qy   1918 CAGGAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCA 1977
          ||     |||||| ||| ||||||||||  | || || || ||||||||||||||    |
Db    320 caaagacatgcattgagctgaagaagcatctcgtcactgctagctatgttgatgatgaaa 379

Qy   1978 TGTCAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCC 2037
           ||| |  ||     |    || ||   ||| ||  |||| || ||  | |||||||| |
Db    380 cgtccatgttacatgagaaggctaagagtgctgggataacgattctaggcgaaatgggac 439

Qy   2038 TAGATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAA 2097
          | || ||||| || |||||| ||||| | ||||| |||||  | || ||||||     ||
Db    440 tggaccctggaatcgatcacatgatggcgatgaaaatgatcaacgatgctcatatcaaaa 499

Qy   2098 AGGGAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAA 2157
          | || ||| | ||| | ||||| ||||| ||||| ||  | || ||||| ||||||||||
Db    500 aagggaaagtgaagtcttttacctcttattgtggagggcttccctctcctgctgcagcaa 559

Qy   2158 ACAATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAA 2217
          | |||||  | || |||||||| || |||||||| ||||| ||| | ||  | ||  |||
Db    560 ataatccattagcatataaatttagctggaaccctgctggagcaattcgagctggtcaaa 619

Qy   2218 ATCCTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATG 2277
          | || ||    ||||||      || || |  || ||||| |||||  | ||  | ||||
Db    620 accccgccaaatacaaaagcaacggcgacataatacatgttgatgggaagaatctctatg 679

Qy   2278 AATCAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAA 2337
          | || ||    ||  || || |||   | ||||||||||||||  ||||    || ||||
Db    680 attccgcggcaagattccgagtacctaatcttccagcttttgcattggagtgttttccaa 739

Qy   2338 ATCGGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATAT 2397
          ||||  | |||||| | || || || | |||||| |||   |  |||||| | || ||||
Db    740 atcgtgactccttggtttacggggaacattatggcatcgagagcgaagcaacaacgatat 799

Qy   2398 ATAGGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTG 2457
           | | |  || :| | || ||||| ||||||  ||| |||| ||| ||||| |||  ||
Db    800 ttcgtggaacactcagatatgaagggtttagtatgataatggcaacactttcgaaacttg 859

Qy   2458 GGTTCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATAAGG 2517
          | ||||||||   || |    ||| | | ||      |   |   |    || | |   |
Db    860 gattctttgacagtgaagcaaatcaagtactctccactggaaagaggattacgtttggtg 919

Qy   2518 GTTTCCTTGATGAACTACTGAATAATATCTCCACAATTAACACGGACTTAGATATTGAAG 2577
           | |  |   | |  | || |||||     |    | | |  |               ||
Db    920 ctcttttaagtaacattctaaataaggatgcagacaatgaatc---------------ag 964

Qy   2578 CTTCTGGTGGATACGATGATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTTGCAAAA 2637
             |    |     || || ||  | |     ||| |  | ||||| ||   || |
Db    965 agcccctagcgggagaagaagagataagcaagagaattatcaagcttggacattcc---- 1020

Qy   2638 ATAAGGAAATAGCTGTTAAGACAGTCAAAACCATCAAGTTCTTGGGACTACATGAAGAGA 2697
            ||||| |  || |  ||  |   |||||| ||    |||||||| |  | |||||||
Db   1021 --aaggagactgcagccaaagctgccaaaacaattgtattcttggggttcaacgaagaga 1078

Qy   2698 CTCAAATACCTAAGGGTTGTTCGAGCCCATTTGATGTGATTTGCCAGCGAATGGAACAGA 2757
             |  | ||       |||    |||  ||||||||  | |||  | | ||||||| |||
Db   1079 gggaggttccatcactgtgtaaaagcgtatttgatgcaacttgttacctaatggaagaga 1138

Qy   2758 GGATGGCCTATGGCCACAATGAGCAAGACATGGTACTGCTCCACCACGAAGTCGAGGTGG 2817
             | || |||  |   ||||| || |||||||| ||  | || |||||||| || ||||
Db   1139 aactagcttattccggaaatgaacaggacatggtgcttttgcatcacgaagtagaagtgg 1198

Qy   2818 AATACCCGGACGGGCAACCCGCCGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGAAGG 2877
          ||| ||  ||  |  |||     || ||||||   ||||| ||  |||| |||||| |
Db   1199 aattccttgaaagcaaacgtatagagaagcacactgcgactcttttggaattcggggaca 1258

Qy   2878 TTGAAAATGGCAGGTCCACCACTGCCATGGCGCTGACCGTCGGCATTCCAGCAGCAATAG 2937
          |  | |||||     | || || || |||||   ||| || || || || ||||| || |
Db   1259 tcaagaatggacaaacaacaaccgctatggccaagactgttgggatccctgcagccattg 1318

Qy   2938 GGGCCCTGCTATTGCTAAAGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGCAAC 2997
          | || ||| | ||  |  |  | ||| || |||| | ||||||  | ||||||||  ||
Db   1319 gagctctggtgttaattgaagacaagatcaagacaagaggagtcttaaggcctctcgaag 1378

Qy   2998 CGGAAATCTACGTTCCAGCATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGA 3057
          | ||  | ||  | ||||| ||||| || ||| |  | |  || || |||||| | ||||
Db   1379 cagaggtgtatttgccagctttggatatattgcaagcatatggtataaagctgatggaga 1438

Qy   3058 AAGTGGA 3064
          | |  ||
Db   1439 aggcaga 1445


RESULT   7
AAV99567
ID   AAV99567 standard; cDNA to mRNA; 429 BP.
XX
AC   AAV99567;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Wheat lysine ketoglutarate reductase partial cDNA.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; wheat; ss.
XX
OS   Triticum aestivum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..252
FT                   /*tag=  a
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   P-PSDB; AAW87765.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 3; Page 199; 231pp; English.
XX
CC   This is the nucleotide sequence of a partial cDNA encoding a
CC   wheat lysine ketoglutarate reductase (LKR) or saccharopine
```

```
CC   dehydrogenase (SDH) polypeptide (see AAW87765).  It was isolated
CC   from an expressed sequence tag database by comparison to an
CC   Arabidopsis LKR/SDH protein (see AAW87759).  Isolated nucleic
CC   acid fragments comprising sequences encoding all or part of plant
CC   LKRs are new.  Also claimed are: (1) a chimeric gene comprising the
CC   fragment (or a subfragment) operably linked to a seed specific
CC   regulator, where the chimeric gene reduces LKR activity in plant
CC   seeds transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence 429 BP; 112 A; 97 C; 127 G; 90 T; 3 other;

  Query Match             5.6%;  Score 183.6;  DB 20;  Length 429;
  Best Local Similarity   69.7%;  Pred. No. 9.7e-42;
  Matches  246;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy   2820 TACCCGGACGGGCAACCCGCCGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGAAGGTT 2879
          ||||| |||||| | ||| |||| ||||||||||||||||| ||||||||||| |||
Db      1 taccccgacggggaccccaccgagaagcaccaagcgacgctgctggagttcggaaagacc 60

Qy   2880 GAAAATGGCAGGTCCACCACTGCCATGGCGCTGACCGTCGGCATTCCAGCAGCAATAGGG 2939
          || || |||||| ||||||| |||||||| || ||||| ||  | || ||||| |||||
Db     61 gagaacggcaggcccaccaccgccatggccctcaccgttggggtaccggcagcgatagga 120

Qy   2940 GCCCTGCTATTGCTAAAGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGCAACCG 2999
          |||||||| |||||  |||| |||||||||| |||||| |||||| ||||| || |||||
Db    121 gccctgctcttgctccagaacaaggtccagaggaaaggggtgatccggcctntggaaccg 180

Qy   3000 GAAATCTACGTTCCAGCATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGAAA 3059
          || || ||| | || ||  |||||||||||||  ||||||||||||||||| | |||| |
Db    181 gagatntacatccctgcgctggagatcttggaagcgtcgggcatcaagctgatcgagaga 240

Qy   3060 GTGGAGACTTGAAAGTTCCCTGATACACAGATAAAGATAGTATGATATAGCAGGGCACAT 3119
          |||||||| |||   |  |  |||     || ||       |  |||||    |   ||
Db    241 gtggagacctgaggatgtcaggatgggatgagaatctatcgagtatatatgctgcagcaa 300

Qy   3120 GTATCTTTTGTATTAACTCCGTTCTGGAATATATATTTGTGAACTAAAATGTG 3172
                     | ||| |    |   || | |     |   || ||||||| |
Db    301 cagaggcagtgagtaaataaaatgatgattntcgccgttgtaagtaaaatgag 353


RESULT    8
ABL12397
ID   ABL12397 standard; cDNA; 2919 BP.
XX
AC   ABL12397;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 31673.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB68294.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 31673; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2919 BP; 686 A; 747 C; 812 G; 674 T; 0 other;

  Query Match             4.8%;  Score 155.8;  DB 23;  Length 2919;
  Best Local Similarity   47.1%;  Pred. No. 2.3e-33;
  Matches  521;  Conservative    0;  Mismatches  572;  Indels   12;  Gaps    1;

Qy     53 GATTATTGTGCAGCCAAGCACAAGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGG 112
          | | || ||||||||   ||   |  |   | | |           ||   | | || ||
Db    294 ggtcatagtgcagccgtccaatcgtcgtgcctaccccatgcaggcctacatgcaggctgg 353

Qy    113 ATGCGAGATTTCAGAAGACCTGTCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAA 172
          |    | |||   || ||| |    ||        | ||  | ||  | |||||
Db    354 agctcatattcaggaggacatcagtgatgcttcggtgatattcggagtgaaacaggtgcc 413

Qy    173 GCTGCAGATGATTCTTTCAGATAGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCA 232
            |  |     |  || | |  |   | |||   |||||||| |||||    ||||| ||
Db    414 catagatgcactgattcctgggaagacctactgcttcttttcgcacacgattaaagcgca 473

Qy    233 AAAAGAGAATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTA 292
            ||   |||||||| || || ||    ||  | ||  | |   |     |  | || ||
Db    474 ggaatcaaatatgccgcttttggatgctattttggagaagaaaattcggctcatcgacta 533

Qy    293 TGAGCTAATTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAG 352
          ||||| |||  | |  ||    ||   | ||    | ||  ||||| |||| ||| ||
Db    534 tgagcgaatcatcgacgaacgcggagcacgacaggtggcctttggcaaatatgccggagt 593

Qy    353 AGCTGGACTGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTC 412
           |||||  || |  |  || | || ||  | |||| |||   ||||  |||  ||| |
Db    594 ggctggcatggtgaacatcctgcacggcattggactgcgtcttttggccctgggacatca 653

Qy    413 GACTCCATTTCTCTCTCTGGGACAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGC 472
           || |||||  |     | ||||   ||||||  ||||       | |    ||
Db    654 tacaccattcatgcacattggacccgctcataactatcgcaattcctcgatggctcgcca 713

Qy    473 TGCAGTCATTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCC 532
           ||| ||   |    ||   | |||||  |           | ||       ||| | ||
Db    714 ggcaatccgggattgtggctacgagatctcgcttggcatgatgcccaagtccattggacc 773

Qy    533 GATAGTGTTTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAA 592
```

```
            |    ||||||||||||||   |||||| || || || ||||| |||||| | |||
Db    774 acttacatttgtgttcactggctctggaaatgtttcccaaggtgcccaggaggtgttctc 833

Qy    593 GTTATTGCCCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAA 652
               |||||  |   | ||||    | |||| ||| | | |  |     |     | ||
Db    834 cgagctgcccattgagtatgttccaccggagatgctacgcaaggtggccgaacatggaaa 893

Qy    653 TCTGTCTAAGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGAC 712
          ||    |||||  |     |       |||   ||   | | |   |   ||  || |
Db    894 tcaaaataagctttacggctgcgaggtgagccgatcggatcatctggagcgtcgtgaggg 953

Qy    713 CTCTAGAGACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTA 772
          |    ||      |      |  | |||     |    ||    |||       || |
Db    954 cggcggatttgatgccaaggaatacgatgagttccccgaacgatacatctccacctttag 1013

Qy    773 TGCTCATCCAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGT 832
            |  |   ||  |  ||| |  | ||| ||  | |  | ||          | |
Db   1014 cacgaagatagctccgtacgcatcggttattgttaacggcatctactgggccgtaggcag 1073

Qy    833 CATCGTAAACTGTATGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTT 892
             |         |    |  |||    | |  ||  | |   |     |    ||  |
Db   1074 tcccaagttgatcagcattccggatgccaagaatctgctacgtccggcaaacactccctg 1133

Qy    893 ACAGCAATTGATGGAGACTGGTTGTCCTTTAG------------TCGGCGTTTGTGACAT 940
           |  ||    ||   ||  | ||    ||||              | | | |||| |||||
Db   1134 gctgccgacgagcaagggtagtccggctttgccacatcgcatgctggccatttgcgacat 1193

Qy    941 AACTTGTGATATTGGAGGTTCCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCC 1000
            |   ||||   || || |||||||| ||||| ||| || | ||  | || || |  ||
Db   1194 ttccgctgatcccggtggctccattgagtttatgaacgagtgcaccaccatcgacactcc 1253

Qy   1001 TTTCTTTCGGTATGATCCTTCTAAGAATTCATACCATGATGATATGGAAGGTGCCGGAGT 1060
           ||||    ||| ||| |    | ||||    |     |     |  | ||  |||||||
Db   1254 attctgcttgtacgatgcggacaggaataaggataccaagagctttaagggacccggagt 1313

Qy   1061 GGTCTGCTTGGCTGTTGACATTCTCCCTACAGAATTCTCTAAAGAGGCCTCCCAACATTT 1120
            |   ||   |  |||| | | | |||||| ||||  | | |||  |  |  |    ||
Db   1314 tttagtctgttcgattgataatatgcctacacaattgcccagagaatcgacggatttgtt 1373

Qy   1121 TGGAAACATACTATCTAGACTTGTT 1145
          |||| |  |  |  ||   | ||||
Db   1374 tggagagctgttggctcctcatgtt 1398


RESULT   9
AAD08977
ID   AAD08977 standard; cDNA; 3266 BP.
XX
AC   AAD08977;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Drosophila melanogaster LKR/SDH cDNA.
XX
KW   Fruitfly; lysine-ketoglutarate reductase; LKR; pesticide; therapy;
KW   saccharopine dehydrogenase; SDH; invertebrate enzyme; drug screening;
KW   peptidyl-alpha-hydroxyglycine alpha-amidating lyase; PAL;
KW   genetic modification; lysine catabolism disorder; ss.
XX
OS   Drosophila melanogaster.
XX
FH   Key             Location/Qualifiers
FT   CDS             104..2890
FT                   /*tag= a
FT                   /product= "dmLKR/SDH protein"
FT   misc_feature    172..1525
FT                   /*tag= b
FT                   /note= "Alanine dehydrogenase/pyridine nucleotide
FT                   transdehyrogenase domain"
FT   misc_feature    1558..1597
FT                   /*tag= c
FT                   /note= "Lactate/malate dehydrogenase domain"
XX
PN   WO200149856-A2.
XX
PD   12-JUL-2001.
XX
PF   05-JAN-2001; 2001WO-US00360.
XX
PR   06-JAN-2000; 2000US-0174973.
PR   29-FEB-2000; 2000US-0185844.
PR   22-MAR-2000; 2000US-0191189.
XX
PA   (GENO-) GENOPTERA LLC.
XX
PI   Stout TJ,  Keegan KP,  Ebens AJ;
XX
DR   WPI; 2001-441885/47.
DR   P-PSDB; AAE05075.
XX
PT   Novel invertebrate enzymes and nucleic acid encoding the enzyme useful
PT   as targets for pesticides and to identify compounds that have utility
PT   as therapeutics or pesticides  -
XX
PS   Claim 1; Page 48-49; 54pp; English.
XX
CC   The invention relates to invertebrate enzymes and nucleic acids,
CC   including lysine-ketoglutarate reductase/saccharopine dehydrogenase
CC   (LKR/SDH) and peptidyl-alpha-hydroxyglycine alpha-amidating lyase (PAL)
CC   from Drosophila melanogaster, designated dmLKR/SDH, dmPAL and dmPAL2.
CC   The enzymes are useful for detecting a candidate compound especially a
CC   putative pesticidal or pharmaceutical agent that interacts with an
CC   invertebrate enzyme or its fragment. The identified compound is useful
CC   for controlling a pest, by reducing viability of the pest. Insect or
CC   worm genetically modified to express the enzyme are useful for studying
CC   invertebrate enzyme activity and for identification and screening of
CC   pesticide targets directed to components of a pathway involving a
CC   subject protein. Nucleic acids encoding the invertebrate enzymes or their
CC   fragments are useful as biopesticides. The enzymes and their DNA are
CC   useful for genetically modifying metazoan invertebrate organisms, such
CC   as insects and worms or cultured cells, resulting in expression or
CC   mis-expression of the protein. The organisms or cultured cells are useful
CC   for screening candidate compounds for the ability to modulate expression
CC   or activity of a subject nucleic acid or protein, and thus are useful in
CC   identification of new drug targets, therapeutic agents, diagnostics and
CC   prognostics useful in treatment of disorders associated with abnormal
CC   enzyme levels and/or function, e.g. disorders associated with lysine
CC   catabolism. The present sequence is Drosophila melanogaster LKR/SDH
CC   (dmLKR/SDH) cDNA. LKR/SDH is a principal enzyme in the catabolism of
CC   lysine.
XX
SQ   Sequence 3266 BP; 812 A; 791 C; 870 G; 793 T; 0 other;


  Query Match             4.8%;  Score 155.8;  DB 22;  Length 3266;
  Best Local Similarity   47.1%;  Pred. No. 2.5e-33;
  Matches  521;  Conservative    0;  Mismatches  572;  Indels   12;  Gaps    1;

Qy     53 GATTATTGTGCAGCCAAGCACAAGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGG 112
          | | || ||||||||   ||   |  |   | | |           ||   | | || ||
Db    265 ggtcatagtgcagccgtccaatcgtcgtgcctaccccatgcaggcctacatgcaggctgg 324

Qy    113 ATGCGAGATTTCAGAAGACCTGTCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAA 172
          |    | |||   || ||| |    ||        | ||  | ||  | |||||
Db    325 agctcatattcaggaggacatcagtgatgcttcggtgatattcggagtgaaacaggtgcc 384

Qy    173 GCTGCAGATGATTCTTTCAGATAGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCA 232
            |  |     |  || | |  |   | |||   |||||||| |||||    ||||| ||
Db    385 catagatgcactgattcctgggaagacctactgcttcttttcgcacacgattaaagcgca 444

Qy    233 AAAAGAGAATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTA 292
            ||   |||||||| || || ||    ||  | ||  | |   |     |  | || ||
Db    445 ggaatcaaatatgccgcttttggatgctattttggagaagaaaattcggctcatcgacta 504

Qy    293 TGAGCTAATTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAG 352
          ||||| |||  | |  ||    ||   | ||    | || ||||| |||| ||| ||
```

```
Db     505 tgagcgaatcatcgacgaacgcggagcacgacaggtggcctttggcaaatatgccggagt 564

Qy     353 AGCTGGACTGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTC 412
            |||||  || |  |  || | || ||  | |||| |||   ||||  ||| ||| |
Db     565 ggctggcatggtgaacatcctgcacggcattggactgcgtcttttggccctgggacatca 624

Qy     413 GACTCCATTTCTCTCTCTGGGACAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGC 472
            || |||||  |     | ||||   ||||||  ||||       | |    ||
Db     625 tacaccattcatgcacattggacccgctcataactatcgcaattcctcgatggctcgcca 684

Qy     473 TGCAGTCATTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCC 532
            ||| ||   |    ||   | |||||  |           | ||       ||| | ||
Db     685 ggcaatccgggattgtggctacgagatctcgcttggcatgatgcccaagtccattggacc 744

Qy     533 GATAGTGTTTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAA 592
             |    ||||||||||||||   |||||| || || || ||||| |||||| | |||
Db     745 acttacatttgtgttcactggctctggaaatgtttcccaaggtgcccaggaggtgttctc 804

Qy     593 GTTATTGCCCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAA 652
                |||||  |   | ||||    | |||| ||| | | |  |     |     | ||
Db     805 cgagctgcccattgagtatgttccaccggagatgctacgcaaggtggccgaacatggaaa 864

Qy     653 TCTGTCTAAGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGAC 712
           ||    |||||  |     |       |||   ||   | | |   |   ||  || |
Db     865 tcaaaataagctttacggctgcgaggtgagccgatcggatcatctggagcgtcgtgaggg 924

Qy     713 CTCTAGAGACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTA 772
           |    ||      |      |  | |||     |    ||    |||       || |
Db     925 cggcggatttgatgccaaggaatacgatgagttccccgaacgatacatctccacctttag 984

Qy     773 TGCTCATCCAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGT 832
             |  |   ||  |  ||| |  | ||| ||  | |  | ||          | |
Db     985 cacgaagatagctccgtacgcatcggttattgttaacggcatctactgggccgtaggcag 1044

Qy     833 CATCGTAAACTGTATGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTT 892
              |         |    |  |||    | |  ||  | |   |     |    ||  |
Db    1045 tcccaagttgatcagcattccggatgccaagaatctgctacgtccggcaaacactccctg 1104

Qy     893 ACAGCAATTGATGGAGACTGGTTGTCCTTTAG------------TCGGCGTTTGTGACAT 940
            | ||    ||   ||  | ||    ||||              | | | |||| |||||
Db    1105 gctgccgacgagcaagggtagtccggctttgccacatcgcatgctggccatttgcgacat 1164

Qy     941 AACTTGTGATATTGGAGGTTCCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCC 1000
             |   ||||   || || |||||||| ||||| ||| || | ||  | || || |  ||
Db    1165 ttccgctgatcccggtggctccattgagtttatgaacgagtgcaccaccatcgacactcc 1224

Qy    1001 TTTCTTTCGGTATGATCCTTCTAAGAATTCATACCATGATGATATGGAAGGTGCCGGAGT 1060
            ||||    ||| ||| |    | ||||    |     |     |  | ||  |||||||
Db    1225 attctgcttgtacgatgcggacaggaataaggataccaagagctttaagggacccggagt 1284

Qy    1061 GGTCTGCTTGGCTGTTGACATTCTCCCTACAGAATTCTCTAAAGAGGCCTCCCAACATTT 1120
             |   ||   |  |||| | | | ||||||| ||||  | | |||  |  |  |    ||
Db    1285 tttagtctgttcgattgataatatgcctacacaattgcccagagaatcgacggatttgtt 1344

Qy    1121 TGGAAACATACTATCTAGACTTGTT 1145
           |||| |  |  |  ||   | ||||
Db    1345 tggagagctgttggctcctcatgtt 1369


RESULT  10
ABL12396/c
ID   ABL12396 standard; cDNA; 6209 BP.
XX
AC   ABL12396;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 31670.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB68293.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 31670; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6209 BP; 1742 A; 1402 C; 1329 G; 1736 T; 0 other;


  Query Match             3.4%;  Score 110.4;  DB 23;  Length 6209;
  Best Local Similarity   49.7%;  Pred. No. 2.9e-20;
  Matches  310;  Conservative    0;  Mismatches  311;  Indels    3;  Gaps    1;

Qy    1868 GGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAGGAGTATG 1927
           ||  || || ||| ||  | |||||||    ||| | ||||      | ||| |    ||
Db    2259 GGCAGATGTAGTAGTTTCCCTGCTGCCCTACAGTCTCCATGGCATGGTGGCACGTTATTG 2200

Qy    1928 CATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCATGTCAAACTT 1987
           | | |    |   |  |||||||| || ||||| ||| |  | ||    || ||    ||
Db    2199 CGTGGCAGAGGGCACCCACATGGTTACCGCAAGTTATCTGAACGACGAGATCTCCGGTTT 2140

Qy    1988 GAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCCTAGATCCTGG 2047
           |  | | |  ||||| |     ||||| || ||  |    ||  ||||| | ||||||||
Db    2139 GCACGAGGAGGCCAAGGCCAAGGGTGTGACCATCATGAACGAGGTGGGCTTGGATCCTGG 2080

Qy    2048 CATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAAAGGGAAAAAT 2107
            || |||||| |  || || || ||   ||  | || |  || |        | |    |
Db    2079 TATCGATCACCTTCTGGCACTGGAGTGCATCCACGAGGTGCAGGACAAGGGAGCAGTTGT 2020

Qy    2108 AAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAAACAATCCGCT 2167
             || | ||       || ||||| ||  ||||  | || |      | || ||| | ||
Db    2019 CGAGTCCTTCGTGAGCTATTGTGGCGGTCTGCCTGCACCGGAGCATTCCAATAATGCCCT 1960

Qy    2168 TGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAAATCCTGCAGT 2227
               ||||| |||  ||||   ||    || | ||| |             |||||
Db    1959 AAGATATAAGTTCTCTTGGTCACCCAGGGGAGTACTTCTCAACACACTTTCCGCTGCAAA 1900

Qy    2228 CTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATGAATCAGCAAA 2287
            ||      ||  ||  | |   |  |  |   |||   |     | |         |||
Db    1899 ATATCTGAGTCAGGGACAAATTGTGGAGATTTCTGGAGGTG---GGGAACTCATGTCAAG 1843

Qy    2288 GAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAAATCGGAATTC 2347
               | |||  |       || ||||  || || |||||    || || ||  || ||||
Db    1842 TCCTCGCAGCTTGGATTTCCTGCCAGGATTCGCCCTGGAGGGTTTCCCCAACAGGGATTC 1783
```

```
Qy    2348 CTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATATATAGGGCTAC 2407
           |   | ||| ||  ||||||| ||  |   ||  || |    |||  |   | | |  ||
Db    1782 CACCAAATACGGCAACCTTTACGGCTTAGGCAGGGATGTGCACACTTTGCTTCGTGGAAC 1723

Qy    2408 TYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTGGGTTCTTTGA 2467
             | |||||  | || |||   ||    ||     |  |          ||  |  | ||
Db    1722 CATACGTTATAAGGGCTTTTCGGAATCGATTAAGCCTATGCAGCTCTTGGGACTTATCGA 1663

Qy    2468 TGCTGCAAATCATCCACTGCTGCA 2491
           | | |     ||| || |||| ||
Db    1662 TCCGGAGCCACATGCATTGCTACA 1639


RESULT  11
AAV35857
ID   AAV35857 standard; cDNA; 323 BP.
XX
AC   AAV35857;
XX
DT   25-AUG-1998  (first entry)
XX
DE   cDNA encoding protein homologous to fungal saccharopine dehydrogenases.
XX
KW   Dihydrodipicolinic acid synthase; DHDPS; chimeric gene;
KW   lysine inhibition; plant chloroplast transit sequence;
KW   plant seed-specific regulatory sequence; transgenic plant;
KW   increased lysine level; corn; Zea mays; soybean; Glycine max; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   US5773691-A.
XX
PD   30-JUN-1998.
XX
PF   07-JUN-1995;   95US-0474633.
XX
PR   19-MAR-1992;   92US-0855414.
PR   18-MAR-1993;   93WO-US02480.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Falco SC,  Keeler SJ,  Rice JA;
XX
DR   WPI; 1998-387117/33.
DR   P-PSDB; AAW60533.
XX
PT   Chimeric genes encoding lysine production enzymes - useful for
PT   increasing transgenic seed lysine content without being inhibited by
PT   high levels of the amino acid
XX
PS   Example 20; Columns 155-156; 106pp; English.
XX
CC   The present sequence encodes a protein homologous to fungal saccharopine
CC   dehydrogenases. The specification describes a Corynebacterium dapA gene,
CC   which encodes a dihydrodipicolinic acid synthase (DHDPS) enzyme, which
CC   was used to create chimeric genes of the invention. The chimeric genes
CC   contain a nucleic acid fragment encoding a DHDPS enzyme which is
CC   insensitive to inhibition by lysine operably linked to a plant
CC   chloroplast transit sequence and to a plant seed-specific regulatory
CC   sequence. The chimeric genes are useful for producing plants containing
CC   increased levels of lysine, especially in corn (Zea mays) and soybean
CC   (Glycine max).
XX
SQ   Sequence 323 BP; 105 A; 57 C; 86 G; 72 T; 3 other;


  Query Match             3.2%;  Score 105.2;  DB 19;  Length 323;
  Best Local Similarity   65.1%;  Pred. No. 1.7e-19;
  Matches  155;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy    2844 AAGCACCAAGCGACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCACCACTGCC 2903
           ||||||   ||||| ||  |||| |||||| |  |  | |||||     | || || ||
```

```
Db       3 aagcacactgcgactcttttggaattcggggacatcaagaatggacaaacaacaaccgct 62

Qy    2904 ATGGCGCTGACCGTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAG 2963
           |||||   ||| || || || || ||||| || || || ||||| ||  |  |  | |||
Db      63 atggccaagactgttgggatccctgcagccattggagctctgctgttaattgaagacaag 122

Qy    2964 GTCCAGACGAAAGGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAGCATTGGAG 3023
            || |||| | ||||||  | ||||||||  || | ||  | ||  | ||||| |||||
Db     123 atcaagacaagaggagtcttaaggcctctcgaagcagaggtgtatttgccagctttggat 182

Qy    3024 ATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGAAAGTTCCCTG 3081
           || ||| |  | |  || || |||||| | ||||| |  ||    | ||||  | |||
Db     183 atattgcaagcatatggtataaagctgatggagaaggcagaatgatcaaagaactctg 240


RESULT  12
AAV99554
ID   AAV99554 standard; DNA; 323 BP.
XX
AC   AAV99554;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Arabidopsis lysine ketoglutarate reductase DNA.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..227
FT                   /*tag=  a
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   P-PSDB; AAW87758.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Example 20; Page 165-166; 231pp; English.
XX
CC   This nucleotide sequence encodes the lysine ketoglutarate reductase
CC   (LKR, see AAW87758), or saccharopine dehydrogenase (SDH), of
CC   Arabidopsis thaliana.  It was isolated on the basis of homology to
CC   fungal glutamate-forming SDH.  Isolated nucleic acid fragments
CC   comprising sequences encoding all or part of plant LKR are new.
CC   Also claimed are: (1) a chimeric gene comprising the fragment (or a
CC   subfragment) operably linked to a suitable seed specific regulator,
CC   where the chimeric gene reduces LKR activity in plant seeds
CC   transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
```

```
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence 323 BP; 104 A; 57 C; 86 G; 72 T; 4 other;


  Query Match             3.2%;  Score 105.2;  DB 20;  Length 323;
  Best Local Similarity   65.1%;  Pred. No. 1.7e-19;
  Matches  155;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy   2844 AAGCACCAAGCGACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCACCACTGCC 2903
          ||||||   ||||| ||  |||| |||||| |  |  | |||||     | || || ||
Db      3 aagcacactgcgactcttttggaattcggggacatcaagaatggacaaacaacaaccgct 62

Qy   2904 ATGGCGCTGACCGTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAG 2963
          |||||   ||| || || || || ||||| || || || ||||| ||  |  |  | |||
Db     63 atggccaagactgttgggatccctgcagccattggagctctgctgttaattgaagacaag 122

Qy   2964 GTCCAGACGAAAGGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAGCATTGGAG 3023
           || |||| | ||||||  | ||||||||  || | ||  | ||  | ||||| |||||
Db    123 atcaagacaagaggagtcttaaggcctctcgaagcagaggtgtatttgccagctttggat 182

Qy   3024 ATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGAAAGTTCCCTG 3081
          || ||| |  | |  || || |||||| | ||||| |  ||    | ||||  | |||
Db    183 atattgcaagcatatggtataaagctgatggagaaggcagaatgatcaaagaactctg 240


RESULT  13
AAV99566
ID   AAV99566 standard; cDNA to mRNA; 308 BP.
XX
AC   AAV99566;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Rice lysine ketoglutarate reductase 3' cDNA.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; rice; ss.
XX
OS   Oryza sativa.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..129
FT                   /*tag=  a
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   P-PSDB; AAW87764.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 3; Page 197; 231pp; English.
```

```
XX
CC   This is the nucleotide sequence of a partial (3') cDNA encoding a
CC   rice lysine ketoglutarate reductase (LKR) or saccharopine
CC   dehydrogenase (SDH) polypeptide (see AAW87764).  It was isolated
CC   from an expressed sequence tag database by comparison to an
CC   Arabidopsis LKR/SDH protein (see AAW87759).  Isolated nucleic
CC   acid fragments comprising sequences encoding all or part of plant
CC   LKRs are new.  Also claimed are: (1) a chimeric gene comprising the
CC   fragment (or a subfragment) operably linked to a seed specific
CC   regulator, where the chimeric gene reduces LKR activity in plant
CC   seeds transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence 308 BP; 96 A; 55 C; 72 G; 85 T; 0 other;


  Query Match             3.1%;  Score 101;  DB 20;  Length 308;
  Best Local Similarity   85.0%;  Pred. No. 2.6e-18;
  Matches  113;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy   2943 CTGCTATTGCTAAAGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGCAACCGGAA 3002
          ||||| |||||  |||| ||| |||| | |||||||||||||||||||||| |||| |||
Db      1 ctgctgttgctccagaacaagatccaaaagaaaggagtgatcaggcctctggaacctgaa 60

Qy   3003 ATCTACGTTCCAGCATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGAAAGTG 3062
          || ||| ||||||| ||||||||||||||||| ||||| ||||||||||  |||| ||||
Db     61 atttacattccagcgttggagatcttggagtcatcgggtatcaagctggcggagagagtg 120

Qy   3063 GAGACTTGAAAGT 3075
          ||||| ||| | |
Db    121 gagacctgagaat 133


RESULT  14
AAF13730
ID   AAF13730 standard; cDNA; 1262 BP.
XX
AC   AAF13730;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Aspergillus oryzae EST SEQ ID NO:6253.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Aspergillus oryzae.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
```

```
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Berka RM,  Rey MW,  Shuster JR,  Kauppinen S,  Clausen IG,  Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 88; Page 2572; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 1262 BP; 259 A; 406 C; 315 G; 279 T; 3 other;


  Query Match             3.0%;  Score 97.6;  DB 21;  Length 1262;
  Best Local Similarity   48.6%;  Pred. No. 5.2e-17;
  Matches  265;  Conservative    0;  Mismatches  280;  Indels    0;  Gaps    0;

Qy    1744 TTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTATTG 1803
           |||| || |  || || |        |   ||   ||   ||||   | ||| ||| |
Db     109 ttcacgtcactgtcgcctgccgtaccctcgaaagcgcccaaaagcttgctgagggtttca 168

Qy    1804 AAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATCTTGTTT 1863
           | || || || || |      | |||||    |||   |     ||     |     |
Db     169 agaacaccacggccattgccttagatgtcaacgatgccgacgctctcgacaaggccctcg 228

Qy    1864 CTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAGGAG 1923
             ||||| ||  |||  ||   |||| | |||   |  || || ||     ||
Db     229 agcaggtcgaccttgctatctccttgatcccttacaccttccacgcccttgttatcaagt 288

Qy    1924 TATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCATGTCAA 1983
                ||       ||||||||  | || ||  |    || ||   |    | |||
Db     289 ccgctatccgcaccaagaagcatgttgttactacttcttacgtctctcncgcaatgctcg 348

Qy    1984 ACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCCTAGATC 2043
           | |||  | |||    |||  | || ||| | ||  |  |    || || ||| | || |
Db     349 agttggacgaagagtgcaagaaggccggtatcacagtcatgaacgagattggcttggacc 408

Qy    2044 CTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAAAGGGAA 2103
           |||| || || ||||||    |  | ||||  ||||| || | ||||     |   || |
Db     409 ctggtattgaccacttgtacgccgtcaagactattgacgaggttcatagggaaggtggta 468

Qy    2104 AAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAAACAATC 2163
           |||| |   | ||    || |||||||||||  | ||  | || |      |  |||| |
Db     469 aaatcacctccttcctctcctactgtggtggtctccccgcccctgaatgctccgacaacc 528
```

```
Qy    2164 CGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAAATCCTG 2223
           | || | ||| || |||   |||        |||||  || |   |       ||  | |
Db     529 ctctgggctacaagttctcctggtctagccgtggtgtccttctcgctctccgtaacgccg 588

Qy    2224 CAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATGAATCAG 2283
           |    | | |       ||  ||    ||   || |  ||||   || |    |   | |
Db     589 ccaagttctacaaggacggccaggaagtcagcgtcgccggtcccgaccttatggctaccg 648

Qy    2284 CAAAG 2288
           | |||
Db     649 ccaag 653


RESULT 15
AAV99559
ID   AAV99559 standard; DNA; 8160 BP.
XX
AC   AAV99559;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Arabidopsis lysine ketoglutarate reductase gene.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; ds.
XX
OS   Arabidopsis thaliana cv. Landsberg erecta.
XX
FH   Key             Location/Qualifiers
FT   CDS             2073..7564
FT                   /*tag=  a
FT                   /note= "contains introns"
FT   exon            2073..2351
FT                   /*tag=  b
FT                   /number= 1
FT   intron          2342..2431
FT                   /*tag=  c
FT                   /number= 1
FT   exon            2432..2518
FT                   /*tag=  d
FT                   /number= 2
FT   intron          2519..2596
FT                   /*tag=  e
FT                   /number= 2
FT   exon            2597..2726
FT                   /*tag=  f
FT                   /number= 3
FT   intron          2727..2809
FT                   /*tag=  g
FT                   /number= 3
FT   exon            2810..2983
FT                   /*tag=  h
FT                   /number= 4
FT   intron          2983..3063
FT                   /*tag=  i
FT                   /number= 4
FT   exon            3064..3147
FT                   /*tag=  j
FT                   /number= 5
FT   intron          3148..3214
FT                   /*tag=  k
FT                   /number= 5
FT   exon            3215..3331
FT                   /*tag=  l
FT                   /number= 6
FT   intron          3332..3424
FT                   /*tag=  m
FT                   /number= 6
FT   exon            3425..3500
FT                   /*tag=  n
FT                   /number= 7
FT   intron          3501..3632
```

```
FT                     /*tag=  o
FT                     /number= 7
FT   exon              3633..3804
FT                     /*tag=  p
FT                     /number= 8
FT   intron            3805..3887
FT                     /*tag=  q
FT                     /number= 8
FT   exon              3888..3984
FT                     /*tag=  r
FT                     /number= 9
FT   intron            3985..4074
FT                     /*tag=  s
FT                     /number= 9
FT   exon              4075..4241
FT                     /*tag=  t
FT                     /number= 10
FT   intron            4242..4329
FT                     /*tag=  u
FT                     /number= 10
FT   exon              4330..4390
FT                     /*tag=  v
FT                     /number= 11
FT   intron            4391..4580
FT                     /*tag=  w
FT                     /number= 11
FT   exon              4581..4721
FT                     /*tag=  x
FT                     /number= 12
FT   intron            4722..4924
FT                     /*tag=  y
FT                     /number= 12
FT   exon              4925..5275
FT                     /*tag=  z
FT                     /number= 13
FT   intron            5276..5369
FT                     /*tag=  aa
FT                     /number= 13
FT   exon              5370..5450
FT                     /*tag=  ab
FT                     /number= 14
FT   intron            5451..5543
FT                     /*tag=  ac
FT                     /number= 14
FT   exon              5544..5609
FT                     /*tag=  ad
FT                     /number= 15
FT   intron            5610..5761
FT                     /*tag=  ae
FT                     /number= 15
FT   exon              5762..5879
FT                     /*tag=  af
FT                     /number= 16
FT   intron            5880..6002
FT                     /*tag=  ag
FT                     /number= 16
FT   exon              6003..6132
FT                     /*tag=  ah
FT                     /number= 17
FT   intron            6133..6233
FT                     /*tag=  ai
FT                     /number= 17
FT   exon              6234..6313
FT                     /*tag=  aj
FT                     /number= 18
FT   intron            6314..6409
FT                     /*tag=  ak
FT                     /number= 18
FT   exon              6410..6568
FT                     /*tag=  al
FT                     /number= 19
FT   intron            6569..6649
FT                     /*tag=  am
```

```
FT                     /number= 19
FT   exon              6650..6881
FT                     /*tag=  an
FT                     /number= 20
FT   intron            6882..6977
FT                     /*tag=  ao
FT                     /number= 20
FT   exon              6978..7086
FT                     /*tag=  ap
FT                     /number= 21
FT   intron            7087..7186
FT                     /*tag=  aq
FT                     /number= 21
FT   exon              7187..7348
FT                     /*tag=  ar
FT                     /number= 22
FT   intron            7349..7442
FT                     /*tag=  as
FT                     /number= 22
FT   exon              7443..7564
FT                     /*tag=  at
FT                     /number= 23
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   P-PSDB; AAW87759.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 3; Page 168-173; 231pp; English.
XX
CC   This is the nucleotide sequence of the Arabidopsis thaliana
CC   lysine ketoglutarate reductase (LKR)/saccharopine dehydrogenase
CC   (SDH) gene that includes exons coding for a 1064-amino acid
CC   protein (see AAW87759).  The gene was isolated from a CD4-8 Landsberg
CC   erecta genomic library using a partial cDNA clone as probe.
CC   Isolated nucleic acid fragments comprising sequences encoding all
CC   or part of plant LKR polypeptides are new.  Also claimed are: (1) a
CC   chimeric gene comprising the fragment (or a subfragment) operably
CC   linked to a suitable seed specific regulator, where the chimeric
CC   gene reduces LKR activity in plant seeds transformed with it; (2)
CC   plant cells and seeds in which LKR activity is reduced due to a
CC   mutation in the gene encoding LKR or transformation with the
CC   chimeric gene; (3) a nucleic acid fragment comprising: (i) chimeric
CC   gene above, and (ii) a second chimeric gene, in which a nucleic
CC   acid fragment encoding dihydrodipicolinic acid synthase (DHDPS)
CC   substantially insensitive to lysine inhibition is operably linked
CC   to a plant chloroplast transit sequence and to a plant
CC   seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,

  Query Match             2.9%;  Score 95.8;  DB 20;  Length 8160;
  Best Local Similarity   78.2%;  Pred. No. 4.9e-16;
```

```
  Matches  115;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy    1325 GGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACATCAT 1384
           |||||||||  || |  |||||||||||||||||| || ||||| |||||| | || || ||
Db    4580 ggtatctctgagcggacacctatttgataagtttctgataaacgaagctcttgatatgat 4639

Qy    1385 TGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGGATGA 1444
            ||  | ||||| || |||||||| |||| || ||||||| | || || ||| | |||| 
Db    4640 cgaagcggctggtggctcatttcatttggctaaatgtgaactggggcagagcgctgatgc 4699

Qy    1445 TATGTCATACTCAGAGCTTGAAGTAGG 1471
           |   || |||||||| ||||||||| |
Db    4700 tgaatcgtactcagaacttgaagtaag 4726


Search completed: June  1, 2002, 02:11:27
Job time: 6987 sec
```

```
                        GenCore version 4.5
               Copyright (c) 1993 - 2000  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          May 31, 2002, 23:20:20 ; Search time 3963.52 Seconds
                                            (without alignments)
                                            17238.501 Million cell updates/sec

Title:           US-09-049-304A-120
Perfect score:   3265
Sequence:        1 ATTGTGCCCGCCTTCTGCTA..........TCAAATAATGCCGATCAGTT 3265

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters:      3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :          GenEmbl:*
                 1:  gb_ba:*
                 2:  gb_htg:*
                 3:  gb_in:*
                 4:  gb_om:*
                 5:  gb_ov:*
                 6:  gb_pat:*
                 7:  gb_ph:*
                 8:  gb_pl:*
                 9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htgo_inv:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                               SUMMARIES
                    %
Result            Query
  No.   Score    Match Length DB  ID                            Description
-------------------------------------------------------------------------------
     1  3264.6  100.0   3265   6  BD010453                BD010453 Chimeric
     2  3179.8   97.4   3518   8  AF003551                AF003551 Zea mays
     3    1908   58.4   1908   6  BD010454                BD010454 Chimeric
     4  1242.2   38.0   3373   8  ATU95759                U95759 Arabidopsis
     5  1236.4   37.9   3195   6  BD010447                BD010447 Chimeric
     6    1232   37.7   3267   8  ATU90522                U90522 Arabidopsis
     7     907   27.8   2582   6  BD010452                BD010452 Chimeric
     8   634.6   19.4   1625   8  AF295389                AF295389 Arabidops
     9   593.2   18.2    620   6  BD010451                BD010451 Chimeric
    10   537.8   16.5   1905   8  AY039906                AY039906 Arabidops
    11   522.6   16.0    720   6  BD010455                BD010455 Chimeric
    12     515   15.8   1449   6  BD010458                BD010458 Chimeric
    13   514.4   15.8   1596   8  ATU90523                U90523 Arabidopsis
    14   326.6   10.0  15146   8  AF271636                AF271636 Zea mays
    15   300.6    9.2    619   6  BD010450                BD010450 Chimeric
    16   238.4    7.3    533   8  AF042184                AF042184 Brassica
    17   220.6    6.8 138504   2  AP004095                AP004095 Oryza sat
    18   183.6    5.6    429   6  BD010457                BD010457 Chimeric
    19     165    5.1   3098   9  AF229180                AF229180 Homo sapi
    20     165    5.1   3233   9  HSA007714               AJ007714 Homo sapi
    21   163.4    5.0   3369  10  BC005420                 BC005420 Mus muscu
    22   161.8    5.0   3300  10  MMA224761                AJ224761 Mus muscu
    23   155.8    4.8   3266   6  AX191629                AX191629 Sequence
    24   152.8    4.7    274   8  AF293461                AF293461 Brassica
c   25   137.4    4.2  33398   8  SPBC3B8                 AL022244 S.pombe c
c   26     114    3.5  42064   3  CBRG44014               AC084621 Caenorhab
c   27   110.4    3.4  76619   3  AC004722                AC004722 Drosophil
    28   110.4    3.4  88165   2  AC014754                AC014754 Drosophil
c   29   110.4    3.4 166722   3  AC008328                AC008328 Drosophil
c   30   110.4    3.4 178863   3  AC093195                AC093195 Drosophil
c   31   110.4    3.4 268667   3  AE003618                AE003618 Drosophil
    32   105.2    3.2    323   6  AR014649                AR014649 Sequence
    33   105.2    3.2    323   6  BD010441                BD010441 Chimeric
    34   101.6    3.1   3240   8  AF144424                AF144424 Magnaport
    35     101    3.1    308   6  BD010456                BD010456 Chimeric
    36    99.8    3.1  32218   3  AF038615                AF038615 Caenorhab
    37    96.8    3.0   1341   8  SCLYS9                  X77363 S.cerevisia
    38    96.8    3.0   1594   8  AF191667                AF191667 Brassica
c   39    96.8    3.0   3124   8  SCYNR050C               Z71665 S.cerevisia
    40    95.8    2.9   8160   6  BD010446                BD010446 Chimeric
    41    95.8    2.9   8160   8  ATU95758                U95758 Arabidopsis
c   42    95.8    2.9 111876   8  ATF4I10                 AL035525 Arabidops
c   43    95.8    2.9 199536   8  ATCHRIV79               AL161583 Arabidops
    44    92.8    2.8    372   6  AR014648                AR014648 Sequence
    45    92.8    2.8    372   6  BD010440                BD010440 Chimeric

                              ALIGNMENTS

RESULT   1
BD010453
LOCUS       BD010453               3265 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Chimeric genes and methods for increasing the lysine content of the
             seeds  of plants Chimeric genes and methods for increasing the
            lysine content of the seeds  of plants.
ACCESSION   BD010453
VERSION     BD010453.1  GI:18638826
KEYWORDS    JP 2001502923-A/85.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3265)
  AUTHORS   Falco,S.C., Iii,R.E.M. and Epelbaum,S.U.
  TITLE     Chimeric genes and methods for increasing the lysine content of the
            seeds of plants
  JOURNAL   Patent: JP 2001502923-A 85 06-MAR-2001;
            EI DU PONT DE NEMOURS AND CO
COMMENT     OS   Zea mays
            PN   JP 2001502923-A/85
            PD   06-MAR-2001
```

```
          PF   27-MAR-1998 JP 1998543284
          PR   27-MAR-1997 US    08/824627
          PI   SAVERIO CARL FALCO,RAYMOND ERVIN MCDEVITT III, PI   SABINE
          URSULA EPELBAUM
          PC   C12N9/06,C12N9/12,C12N9/88,C12P13/08,C12N15/82 CC
          Strandedness: Single;
          CC   Topology: Linear;
          FH   Key              Location/Qualifiers
          FT   CDS              3. .3071.
FEATURES             Location/Qualifiers
     source          1. .3265
                     /organism="Zea mays"
                     /db_xref="taxon:4577"
BASE COUNT      952 a    651 c    765 g    896 t      1 others
ORIGIN

  Query Match             100.0%;  Score 3264.6;  DB 6;  Length 3265;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3265;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTGTGCCCGCCTTCTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATTGTGCCCGCCTTCTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTG 60

Qy         61 TGCAGCCAAGCACAAGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGCAGCCAAGCACAAGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGA 120

Qy        121 TTTCAGAAGACCTGTCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTTCAGAAGACCTGTCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGA 180

Qy        181 TGATTCTTTCAGATAGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGATTCTTTCAGATAGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGA 240

Qy        241 ATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAA 300

Qy        301 TTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGAC 360

Qy        361 TGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCAT 420

Qy        421 TTCTCTCTCTGGGACAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTCTCTCTCTGGGACAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCA 480

Qy        481 TTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATAGTGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATAGTGT 540

Qy        541 TTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGC 600

Qy        601 CCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTA 660

Qy        661 AGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAG 720

Qy        721 ACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATC 780

Qy        781 CAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAA 840

Qy        841 ACTGTATGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ACTGTATGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAAT 900

Qy        901 TGATGGAGACTGGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGATGGAGACTGGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTT 960

Qy        961 CCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTT 1020

Qy       1021 CTAAGAATTCATACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTAAGAATTCATACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACA 1080

Qy       1081 TTCTCCCTACAGAATTCTCTAAAGAGGCCTCCCAACATTTGGAAACATACTATCTAGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TTCTCCCTACAGAATTCTCTAAAGAGGCCTCCCAACATTTGGAAACATACTATCTAGAC 1140

Qy       1141 TTGTTGCTAGTTTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTGTTGCTAGTTTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAG 1200

Qy       1201 CTTGCATTGCACATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CTTGCATTGCACATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAA 1260

Qy       1261 ATACTATGATAGATTTGGCACCCGCAAAAACAAATCCATTGCCTGACAAGAAGTATAGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATACTATGATAGATTTGGCACCCGCAAAAACAAATCCATTGCCTGACAAGAAGTATAGCA 1320

Qy       1321 CCCTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACA 1380

Qy       1381 TCATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGG 1440

Qy       1441 ATGATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATGATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAA 1500

Qy       1501 TTATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCAAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TTATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCAAG 1560

Qy       1561 AAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAATTG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 AAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAATTG 1620

Qy       1621 ATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTG 1680

Qy       1681 AGTTTCTGGCATCTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 AGTTTCTGGCATCTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATC 1740

Qy       1741 AAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTA 1800

Qy       1801 TTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATCTTG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATCTTG 1860

Qy       1861 TTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAG 1920
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 TTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAG 1920

Qy     1921 GAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCATGT 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 GAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCATGT 1980

Qy     1981 CAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCCTAG 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 CAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCCTAG 2040

Qy     2041 ATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAAAGG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 ATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAAAGG 2100

Qy     2101 GAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAAACA 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 GAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAAACA 2160

Qy     2161 ATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAAATC 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2161 ATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAAATC 2220

Qy     2221 CTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATGAAT 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2221 CTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATGAAT 2280

Qy     2281 CAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAAATC 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2281 CAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAAATC 2340

Qy     2341 GGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATATATA 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2341 GGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATATATA 2400

Qy     2401 GGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTGGGT 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2401 GGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTGGGT 2460

Qy     2461 TCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATAAGGGTT 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2461 TCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATAAGGGTT 2520

Qy     2521 TCCTTGATGAACTACTGAATAATATCTCCACAATTAACACGGACTTAGATATTGAAGCTT 2580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2521 TCCTTGATGAACTACTGAATAATATCTCCACAATTAACACGGACTTAGATATTGAAGCTT 2580

Qy     2581 CTGGTGGATACGATGATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTTGCAAAAATA 2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2581 CTGGTGGATACGATGATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTTGCAAAAATA 2640

Qy     2641 AGGAAATAGCTGTTAAGACAGTCAAAACCATCAAGTTCTTGGGACTACATGAAGAGACTC 2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2641 AGGAAATAGCTGTTAAGACAGTCAAAACCATCAAGTTCTTGGGACTACATGAAGAGACTC 2700

Qy     2701 AAATACCTAAGGGTTGTTCGAGCCCATTTGATGTGATTTGCCAGCGAATGGAACAGAGGA 2760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2701 AAATACCTAAGGGTTGTTCGAGCCCATTTGATGTGATTTGCCAGCGAATGGAACAGAGGA 2760

Qy     2761 TGGCCTATGGCCACAATGAGCAAGACATGGTACTGCTCCACCACGAAGTCGAGGTGGAAT 2820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2761 TGGCCTATGGCCACAATGAGCAAGACATGGTACTGCTCCACCACGAAGTCGAGGTGGAAT 2820

Qy     2821 ACCCGGACGGGCAACCCGCCGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGAAGGTTG 2880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2821 ACCCGGACGGGCAACCCGCCGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGAAGGTTG 2880

Qy     2881 AAAATGGCAGGTCCACCACTGCCATGGCGCTGACCGTCGGCATTCCAGCAGCAATAGGGG 2940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2881 AAAATGGCAGGTCCACCACTGCCATGGCGCTGACCGTCGGCATTCCAGCAGCAATAGGGG 2940

Qy     2941 CCCTGCTATTGCTAAAGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGCAACCGG 3000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2941 CCCTGCTATTGCTAAAGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGCAACCGG 3000

Qy     3001 AAATCTACGTTCCAGCATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGAAAG 3060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3001 AAATCTACGTTCCAGCATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGAAAG 3060

Qy     3061 TGGAGACTTGAAAGTTCCCTGATACACAGATAAAGATAGTATGATATAGCAGGGCACATG 3120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3061 TGGAGACTTGAAAGTTCCCTGATACACAGATAAAGATAGTATGATATAGCAGGGCACATG 3120

Qy     3121 TATCTTTTGTATTAACTCCGTTCTGGAATATATATTTGTGAACTAAAATGTGACAAATAA 3180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3121 TATCTTTTGTATTAACTCCGTTCTGGAATATATATTTGTGAACTAAAATGTGACAAATAA 3180

Qy     3181 AAAGAACGGGTGGAGTATATTGTAAGAGACGGCAAAGAAACCTCTGTATATATGACCTGT 3240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3181 AAAGAACGGGTGGAGTATATTGTAAGAGACGGCAAAGAAACCTCTGTATATATGACCTGT 3240

Qy     3241 CGATATCAAATAATGCCGATCAGTT 3265
            |||||||||||||||||||||||||
Db     3241 CGATATCAAATAATGCCGATCAGTT 3265


RESULT    2
AF003551
LOCUS       AF003551                3518 bp    mRNA    linear   PLN 22-DEC-1999
DEFINITION  Zea mays lysine-ketoglutarate reductase/saccharopine dehydrogenase
            bifunctional enzyme mRNA, complete cds.
ACCESSION   AF003551
VERSION     AF003551.2  GI:5016094
KEYWORDS    .
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3518)
  AUTHORS   Papes,F., Kemper,E.L., Cord-Neto,G., Langone,F. and Arruda,P.
  TITLE     Lysine degradation through the saccharopine pathway in mammals:
            involvement of both bifunctional and monofunctional
            lysine-degrading enzymes in mouse
  JOURNAL   Biochem. J. 344 Pt 2, 555-563 (1999)
  MEDLINE   20036337
REFERENCE   2  (bases 1 to 3518)
  AUTHORS   Cord-Neto,G., Kemper,E.L. and Arruda,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-1997) Centro de Biologia Molecular e Engenharia
            Genetica, Universidade Estadual de Campinas, Cidade Universitaria,
            Campinas, SP 13083-970, Brazil
REFERENCE   3  (bases 1 to 3518)
  AUTHORS   Cord-Neto,G., Kemper,E.L. and Arruda,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-1998) Centro de Biologia Molecular e Engenharia
            Genetica, Universidade Estadual de Campinas, Cidade Universitaria,
            Campinas, SP 13083-970, Brazil
  REMARK    Sequence update by submitter
REFERENCE   4  (bases 1 to 3518)
  AUTHORS   Cord-Neto,G., Kemper,E.L. and Arruda,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-1999) Centro de Biologia Molecular e Engenharia
            Genetica, Universidade Estadual de Campinas, Cidade Universitaria,
            Campinas, SP 13083-970, Brazil
  REMARK    Sequence update by submitter
COMMENT     On Jun 8, 1999 this sequence version replaced gi:3157909.
FEATURES             Location/Qualifiers
     source          1..3518
                     /organism="Zea mays"
                     /cultivar="AGROCERES F-352 commercial hybrid"
                     /db_xref="taxon:4577"
     CDS             178..3360
                     /codon_start=1
                     /product="lysine-ketoglutarate reductase/saccharopine
                     dehydrogenase bifunctional enzyme"
```

```
                     /protein_id="AAC18622.2"
                     /db_xref="GI:5016095"
                     /translation="MGSAATEGNDTLLGNGVVGILAETCNMWERRAPLTPSHCARLLL
                     GGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQMILSD
                     RAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLID
                     FLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF
                     VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTS
                     RDMVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQ
                     LQQLMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVV
                     CLAVDILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLY
                     EYIPRMRNTMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHL
                     VRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIG
                     KVNEYETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVA
                     SLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIE
                     LKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKI
                     KAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYES
                     AKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATLRYEGFSEIMVTLSKT
                     GFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGC
                     CKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHH
                     EVEVEYPDGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKG
                     VIRPLQPEIYVPALEILESSGIKLVEKVET"
BASE COUNT      998 a    735 c    833 g    952 t
ORIGIN


  Query Match             97.4%;  Score 3179.8;  DB 8;  Length 3518;
  Best Local Similarity   99.3%;  Pred. No. 0;
  Matches 3206;  Conservative    1;  Mismatches   13;  Indels    9;  Gaps    1;

Qy          1 ATTGTGCCCGCCTTCTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 ATTGTGCCCGCCTTCTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTG 349

Qy         61 TGCAGCCAAGCACAAGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGA 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        350 TGCAGCCAAGCACAAGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGA 409

Qy        121 TTTCAGAAGACCTGTCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGA 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        410 TTTCAGAAGACCTGTCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGA 469

Qy        181 TGATTCTTTCAGATAGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGA 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        470 TGATTCTTTCAGATAGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGA 529

Qy        241 ATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAA 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        530 ATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAA 589

Qy        301 TTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGAC 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        590 TTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGAC 649

Qy        361 TGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCAT 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        650 TGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCAT 709

Qy        421 TTCTCTCTCTGGGACAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCA 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        710 TTCTCTCTCTGGGACAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCA 769

Qy        481 TTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATAGTGT 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        770 TTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATAGTGT 829

Qy        541 TTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGC 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        830 TTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGC 889

Qy        601 CCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTA 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        890 CCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTA 949

Qy        661 AGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAG 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        950 AGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAG 1009

Qy        721 ACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATC 780
                |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1010 ACATGGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATC 1069

Qy        781 CAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAA 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1070 CAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAA 1129

Qy        841 ACTGTATGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAAT 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1130 ACTGTATGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAAT 1189

Qy        901 TGATGGAGACTGGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTT 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1190 TGATGGAGACTGGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTT 1249

Qy        961 CCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTT 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1250 CCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTT 1309

Qy       1021 CTAAGAATTCATACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACA 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1310 CTAAGAATTCATACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACA 1369

Qy       1081 TTCTCCCTACAGAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGAC 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1370 TTCTCCCTACAGAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGAC 1429

Qy       1141 TTGTTGCTAGTTTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAG 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1430 TTGTTGCTAGTTTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAG 1489

Qy       1201 CTTGCATTGCACATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAA 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1490 CTTGCATTGCACATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAA 1549

Qy       1261 ATACTATGATAGATTTGGCACCCGCAAAAACAAATCCATTGCCTGACAAGAAGTATAGCA 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1550 ATACTATGATAGATTTGGCACCCGCAAAAACAAATCCATTGCCTGACAAGAAGTATAGCA 1609

Qy       1321 CCCTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACA 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1610 CCCTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACA 1669

Qy       1381 TCATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGG 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1670 TCATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGG 1729

Qy       1441 ATGATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAA 1500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1730 ATGATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAA 1789

Qy       1501 TTATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCAAG 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1790 TTATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCAAG 1849

Qy       1561 AAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAATTG 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1850 AAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAATTG 1909

Qy       1621 ATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTG 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1910 ATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTG 1969

Qy       1681 AGTTTCTGGCATCTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATC 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1970 AGTTTCTGGCATCTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATC 2029

Qy       1741 AAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTA 1800
```

```
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2030 AAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTA 2089

Qy    1801 TTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATCTTG 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2090 TTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATCTTG 2149

Qy    1861 TTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAG 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2150 TTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAG 2209

Qy    1921 GAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCATGT 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2210 GAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCATGT 2269

Qy    1981 CAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCCTAG 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2270 CAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCCTAG 2329

Qy    2041 ATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAAAGG 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2330 ATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAAAGG 2389

Qy    2101 GAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAAACA 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2390 GAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAAACA 2449

Qy    2161 ATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAAATC 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2450 ATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAAATC 2509

Qy    2221 CTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATGAAT 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2510 CTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATGAAT 2569

Qy    2281 CAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAAATC 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2570 CAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAAATC 2629

Qy    2341 GGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATATATA 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2630 GGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATATATA 2689

Qy    2401 GGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTGGGT 2460
           ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2690 GGGCTACTCTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTGGGT 2749

Qy    2461 TCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATAAGGGTT 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2750 TCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATAAGGGTT 2809

Qy    2521 TCCTTGATGAACTACTGAATAATATCTCCACAATTAACACGGACTTAGATATTGAAGCTT 2580
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    2810 TCCTTGATGAACTACTGAATAATATCTCCACAATTAACACGGACTTAGATATTGAAGCCT 2869

Qy    2581 CTGGTGGATACGATGATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTTGCAAAAATA 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2870 CTGGTGGATACGATGATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTTGCAAAAATA 2929

Qy    2641 AGGAAATAGCTGTTAAGACAGTCAAAACCATCAAGTTCTTGGGACTACATGAAGAGACTC 2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2930 AGGAAATAGCTGTTAAGACAGTCAAAACCATCAAGTTCTTGGGACTACATGAAGAGACTC 2989

Qy    2701 AAATACCTAAGGGTTGTTCGAGCCCATTTGATGTGATTTGCCAGCGAATGGAACAGAGGA 2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2990 AAATACCTAAGGGTTGTTCGAGCCCATTTGATGTGATTTGCCAGCGAATGGAACAGAGGA 3049

Qy    2761 TGGCCTATGGCCACAATGAGCAAGACATGGTACTGCTCCACCACGAAGTCGAGGTGGAAT 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3050 TGGCCTATGGCCACAATGAGCAAGACATGGTACTGCTCCACCACGAAGTCGAGGTGGAAT 3109

Qy    2821 ACCCGGACGGGCAACCCGCCGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGAAGGTTG 2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3110 ACCCGGACGGGCAACCCGCCGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGAAGGTTG 3169

Qy    2881 AAAATGGCAGGTCCACCACTGCCATGGCGCTGACCGTCGGCATTCCAGCAGCAATAGGGG 2940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3170 AAAATGGCAGGTCCACCACTGCCATGGCGCTGACCGTCGGCATTCCAGCAGCAATAGGGG 3229

Qy    2941 CCCTGCTATTGCTAAAGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGCAACCGG 3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3230 CCCTGCTATTGCTAAAGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGCAACCGG 3289

Qy    3001 AAATCTACGTTCCAGCATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGAAAG 3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3290 AAATCTACGTTCCAGCATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGAAAG 3349

Qy    3061 TGGAGACTTGAAAGTTCCCTGATACACAGATAAAGATAGTATGATATAGCAGGGCACATG 3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3350 TGGAGACTTGAAAGTTCCCTGATACACAGATAAAGATAGTATGATATAGCAGGGCACATG 3409

Qy    3121 TATCTTTTGTATTAACTCCGTTCTGGAATATATATTTGTGAACTAAAATGTGAC------ 3174
           |||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    3410 TATCTTTTGTATTAACTCCGTTCTGGAATATATATTTGTGAACTAAAACGTGACAAATAA 3469

Qy    3175 ---AAATAAAAAGAACGGGTGGAGTATATTGTAAGAGACGGCAAAGAAA 3220
              |||||||||||||||  | |||||||||||| | |    ||| |||
Db    3470 AAGAAATAAAAAGAACGGAAGTAGTATATTGTAAAAAAAAAAAAAAAAA 3518
```

```
RESULT   3
BD010454
LOCUS       BD010454                1908 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Chimeric genes and methods for increasing the lysine content of the
             seeds  of plants Chimeric genes and methods for increasing the
            lysine content of the seeds  of plants.
ACCESSION   BD010454
VERSION     BD010454.1  GI:18638827
KEYWORDS    JP 2001502923-A/86.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1908)
  AUTHORS   Falco,S.C., Iii,R.E.M. and Epelbaum,S.U.
  TITLE     Chimeric genes and methods for increasing the lysine content of the
            seeds of plants
  JOURNAL   Patent: JP 2001502923-A 86 06-MAR-2001;
            EI DU PONT DE NEMOURS AND CO
COMMENT     OS   Zea mays
            PN   JP 2001502923-A/86
            PD   06-MAR-2001
            PF   27-MAR-1998 JP 1998543284
            PR   27-MAR-1997 US    08/824627
            PI   SAVERIO CARL FALCO,RAYMOND ERVIN MCDEVITT III, PI   SABINE
            URSULA EPELBAUM
            PC   C12N9/06,C12N9/12,C12N9/88,C12P13/08,C12N15/82 CC
            Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   CDS             3. .1908.
FEATURES             Location/Qualifiers
     source          1. .1908
                     /organism="Zea mays"
                     /db_xref="taxon:4577"
BASE COUNT      537 a    376 c    439 g    556 t
ORIGIN


  Query Match              58.4%;  Score 1908;  DB 6;  Length 1908;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1908;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTGTGCCCGCCTTCTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 ATTGTGCCCGCCTTCTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTG 60

Qy         61 TGCAGCCAAGCACAAGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGCAGCCAAGCACAAGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGA 120

Qy        121 TTTCAGAAGACCTGTCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTTCAGAAGACCTGTCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGA 180

Qy        181 TGATTCTTTCAGATAGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGATTCTTTCAGATAGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGA 240

Qy        241 ATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAA 300

Qy        301 TTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGAC 360

Qy        361 TGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCAT 420

Qy        421 TTCTCTCTCTGGGACAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTCTCTCTCTGGGACAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCA 480

Qy        481 TTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATAGTGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATAGTGT 540

Qy        541 TTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGC 600

Qy        601 CCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTA 660

Qy        661 AGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAG 720

Qy        721 ACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATC 780

Qy        781 CAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAA 840

Qy        841 ACTGTATGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ACTGTATGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAAT 900

Qy        901 TGATGGAGACTGGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGATGGAGACTGGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTT 960

Qy        961 CCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTT 1020

Qy       1021 CTAAGAATTCATACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTAAGAATTCATACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACA 1080

Qy       1081 TTCTCCCTACAGAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TTCTCCCTACAGAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGAC 1140

Qy       1141 TTGTTGCTAGTTTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTGTTGCTAGTTTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAG 1200

Qy       1201 CTTGCATTGCACATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CTTGCATTGCACATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAA 1260

Qy       1261 ATACTATGATAGATTTGGCACCCGCAAAAACAAATCCATTGCCTGACAAGAAGTATAGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATACTATGATAGATTTGGCACCCGCAAAAACAAATCCATTGCCTGACAAGAAGTATAGCA 1320

Qy       1321 CCCTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACA 1380

Qy       1381 TCATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGG 1440

Qy       1441 ATGATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATGATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAA 1500

Qy       1501 TTATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCAAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TTATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCAAG 1560

Qy       1561 AAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAATTG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 AAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAATTG 1620

Qy       1621 ATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTG 1680

Qy       1681 AGTTTCTGGCATCTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 AGTTTCTGGCATCTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATC 1740

Qy       1741 AAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTA 1800

Qy       1801 TTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATCTTG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATCTTG 1860

Qy       1861 TTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATG 1908
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATG 1908
```

```
RESULT     4
ATU95759
LOCUS       ATU95759                3373 bp    mRNA    linear   PLN 16-JAN-1998
DEFINITION  Arabidopsis thaliana lysine-ketoglutarate reductase/saccharopine
            dehydrogenase bifunctional enzyme mRNA, complete cds.
ACCESSION   U95759
VERSION     U95759.1  GI:2052507
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 3373)
  AUTHORS   Epelbaum,S., McDevitt,R. and Falco,S.C.
  TITLE     Lysine-ketoglutarate reductase and saccharopine dehydrogenase from
            Arabidopsis thaliana: nucleotide sequence and characterization
  JOURNAL   Plant Mol. Biol. 35 (6), 735-748 (1997)
  MEDLINE   98088001
```

```
REFERENCE   2  (bases 1 to 3373)
  AUTHORS   Epelbaum,S., McDevitt,R. and Falco,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1997) Agricultural Products, E.I. Dupont de
            Nemours & Co., P.O. Box 80402, Wilmington, DE 19880-0402, USA
FEATURES             Location/Qualifiers
     source          1. .3373
                     /organism="Arabidopsis thaliana"
                     /strain="ssp. columbia"
                     /db_xref="taxon:3702"
     CDS             92. .3286
                     /function="lysine catabolism"
                     /codon_start=1
                     /product="lysine-ketoglutarate reductase/saccharopine
                     dehydrogenase bifunctional enzyme"
                     /protein_id="AAB96826.1"
                     /db_xref="GI:2052508"
                     /translation="MNSNGHEEEKKLGNGVVGILAETVNKWERRTPLTPSHCARLLHG
                     GKDRTGISRIVVQPSAKRIHHDALYEDVGCEISDDLSDCGLILGIKQPELEMILPERA
                     YAFFSHTHKAQKENMPLLDKILSERVTLCDYELIVGDHGKRLLAFGKYAGRAGLVDFL
                     HGLGQRYLSLGYSTPFLSLGASYMYSSLAAAKAAVISVGEEIASQGLPLGICPLVFVF
                     TGTGNVSLGAQEIFKLLPHTFVEPSKLPELFVKDKGISQNGISTKRVYQVYGCIITSQ
                     DMVEHKDPSKSFDKADYYAHPEHYNPVFHEKISPYTSVLVNCMYWEKRFPCLLSTKQL
                     QDLTKKGLPLVGICDITCDIGGSIEFVNRATLIDSPFFRFNPSNNSYYDDMDGDGVLC
                     MAVDILPTEFAKEASQHFGDILSGFVGSLASMTEISDLPAHLKRACISYRGELTSLYE
                     YIPRMRKSNPEEAQDNIIANGVSSQRTFNILVSLSGHLFDKFLINEALDMIEAAGGSF
                     HLAKCELGQSADAESYSELEVGADDKRVLDQIIDSLTRLANPNEDYISPHREANKISL
                     KIGKVQQENEIKEKPEMTKKSGVLILGAGRVCRPAADFLASVRTISSQQWYKTYFGAD
                     SEEKTDVHVIVASLYLKDAKETVEGISDVEAVRLDVSDSESLLKYVSQVDVVLSLLPA
                     SCHAVVAKTCIELKKHLVTASYVDDETSMLHEKAKSAGITILGEMGLDPGIDHMMAMK
                     MINDAHIKKGKVKSFTSYCGGLPSPAAANNPLAYKFSWNPAGAIRAGQNPAKYKSNGD
                     IIHVDGKNLYDSAARFRVPNLPAFALECLPNRDSLVYGEHYGIESEATTIFRGTLRYE
                     GFSMIMATLSKLGFFDSEANQVLSTGKRITFGALLSNILNKDADNESEPLAGEEEISK
                     RIIKLGHSKETAAKAAKTIVFLGFNEEREVPSLCKSVFDATCYLMEEKLAYSGNEQDM
                     VLLHHEVEVEFLESKRIEKHTATLLEFGDIKNGQTTTAMAKTVGIPAAIGALLLIEDK
                     IKTRGVLRPLEAEVYLPALDILQAYGIKLMEKAE"
BASE COUNT     1040 a    637 c    778 g    918 t
ORIGIN

  Query Match              38.0%;  Score 1242.2;  DB 8;  Length 3373;
  Best Local Similarity    64.6%;  Pred. No. 3e-309;
  Matches 2009;  Conservative    1;  Mismatches 1034;  Indels   67;  Gaps    8;

Qy      15 CTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTGTGCAGCCAAGCACA 74
           ||  ||   || || ||  || || |  |  |   ||| ||| | || ||||||    |
Db     212 CTTTTACACGGTGGGAAAGACAGAACCGGCATTTCCCGCATTGTGGTTCAGCCATCTGCT 271

Qy      75 AGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGATTTCAGAAGACCTG 134
           | | | ||||||||||| ||   ||||||| ||||  || || || ||||| || ||  ||
Db     272 AAGCGTATCCATCATGATGCCTTGTATGAAGATGTTGGGTGTGAAATTTCTGATGATTTG 331

Qy     135 TCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGATGATTCTTTCAGAT 194
           || || || || |||||  | || ||||||||||||  ||||  | ||||||||| ||||
Db     332 TCTGATTGTGGGCTTATACTTGGAATCAAACAACCTGAGCTAGAAATGATTCTTCCAGAG 391

Qy     195 AGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGAATATGCCACTGTTA 254
           ||||| |||||||||||||||| || || || || || ||||||||| ||||||  ||||
Db     392 AGAGCATACGCTTTCTTTTCACATACTCATAAGGCACAGAAAGAGAACATGCCTTTGTTG 451

Qy     255 GACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAATTGTTGGAGATGAT 314
           || || || |||   || || ||| | |||| |||||||||||| |||||||| ||| ||
Db     452 GATAAAATTCTTTCTGAGAGAGTGACTTTGTGTGATTATGAGCTCATTGTTGGGGATCAT 511

Qy     315 GGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGACTGATAGATTTCTTA 374
           |||||| ||| | | || ||||| |||| ||| || |||||||| ||  | || ||||||
Db     512 GGGAAACGATTATTGGCGTTTGGTAAATATGCAGGCAGAGCTGGTCTTGTTGACTTCTTA 571

Qy     375 CATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCATTTCTCTCTCTGGGA 434
           || || || ||||||||||| | || || |||||||| || || || ||||| || || |
Db     572 CACGGACTTGGACAGCGATATCTAAGTCTAGGATACTCAACACCTTTCCTCTCGCTCGGT 631

Qy     435 CAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCATTGTCGTTGCAGAA 494
             |||  |||||||| | ||  | ||||| || || || || || |||    ||||  |||
Db     632 GCATCGTATATGTATTCCTCATTGGCTGCTGCAAAAGCCGCTGTAATTTCTGTTGGTGAA 691

Qy     495 GAGATAGCAACATTTGGACTTGCATCCGGAATTTGTCCGATAGTGTTTGTGTTCACTGGA 554
           || || ||||     |||||  |||  |||||| || ||  | || ||||| ||||| |||
Db     692 GAAATTGCAAGCCAGGGACTGCCATTAGGAATCTGCCCTCTTGTATTTGTCTTCACCGGA 751

Qy     555 GTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGCCCCATACCTTTGTT 614
              ||||| || |||| ||| ||||| || || |||||| |  | || || || ||||||
Db     752 ACAGGAAATGTTTCTCTGGGGGCGCAAGAAATTTTCAAGCTTCTTCCTCACACTTTTGTT 811

Qy     615 GATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTAAGCAATCTCAG--- 671
           ||  |    || ||||| ||| | |||         ||      ||  |||  |  |
Db     812 GAACCAAGCAAACTTCCTGAACTATTTGTAAAAGACAAAGGAATTAGTCAAAATGGGATT 871

Qy     672 TCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAGACATAGTTTCT 731
           || || ||| |||| | ||| | ||||||||| || | |||     |||||| |||
Db     872 TCAACAAAGCGAGTCTATCAAGTATATGGTTGTATTATTACCAGCCAAGACATGGTTGAA 931

Qy     732 CACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATCCAGAACACTAC 791
           ||||| |||||  | |    ||| |||||||  ||||||||||||| || || ||||| |||
Db     932 CACAAAGATCCATCAAAGTCATTCGACAAAGCCGACTATTATGCACACCCGGAACATTAC 991

Qy     792 ACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAAACTGTATGTAT 851
           |  || ||||| || ||||  ||  | |||||| | |||||  | |||||||||||||||
Db     992 AATCCAGTTTTCCACGAAAAGATATCGCCATATACGTCTGTTCTTGTAAACTGTATGTAC 1051

Qy     852 TGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAATTGATGGAGACT 911
           |||||||||||||||||     | || |  |   | ||| | ||  | || |   | |
Db    1052 TGGGAGAAGAGGTTTCCCTGTCTTCTGAGCACAAAACAGCTTCAAGATTTAACAAAAAAA 1111

Qy     912 GGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTTCCATTGAATTT 971
           ||    ||  |||| ||| | ||||| |||||||||| || || || ||||||||||||
Db    1112 GGACTCCCACTAGTAGGCATATGTGATATAACTTGTGACATCGGTGGCTCCATTGAATTT 1171

Qy     972 ATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTTCTAAGAATTCA 1031
            | |||     ||| | ||| ||    ||||||||  ||| | |||| || || ||||||
Db    1172 GTTAACCGAGCTACTTTAATCGATTCCCCTTTCTTCAGGTTTAATCCCTCGAACAATTCA 1231

Qy    1032 TACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACATTCTCCCTACA 1091
           ||| | ||||| ||||| || |  || ||  | ||| ||||||||||||| | || |||
Db    1232 TACTACGATGACATGGATGGGGATGGCGTACTATGCATGGCTGTTGACATTTTACCCACA 1291

Qy    1092 GAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGACTTGTTGCTAGT 1151
           |||||  | ||||||||| |||| |||||||||| | || || ||  || ||||| | ||||
Db    1292 GAATTTGCAAAAGAGGCATCCCAGCATTTTGGAGATATTCTTTCCGGATTTGTCGGTAGT 1351

Qy    1152 TTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAGCTTGCATTGCA 1211
           ||||| ||| |||   ||    |||| || ||  |  |  |||  || ||||||||
Db    1352 TTGGCTTCAATGACTGAAATTTCAGATCTACCAGCACATCTGAAGAGGGCTTGCATAAGC 1411

Qy    1212 CATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAAATACTATGATA 1271
            ||   ||   ||| ||  ||||||||| ||||| ||  | ||||| ||  | |    |
Db    1412 TATAGGGGAGAATTGACATCTTTGTATGAGTATATTCCACGTATGAGGAAGTCAAATCCA 1471

Qy    1272 GATTTGGCACCCGCAAAAACAAATCCATTGCCTG---------ACAAGAAGTATAGCACC 1322
           ||   |||||  |  || |  |   |       |         | |  |  |  | ||
Db    1472 GAAGAGGCACAAGATAATATTATCGCCAACGGGGTTTCCAGCCAGAGAACATTCAACATA 1531

Qy    1323 CTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACATC 1382
            ||||||||||| || || |||||||||||||||| || ||||| |||||| | || ||
Db    1532 TTGGTATCTCTGAGCGGACACCTATTTGATAAGTTTCTGATAAACGAAGCTCTTGATATG 1591

Qy    1383 ATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGGAT 1442
           || ||  | ||||| || |||||| ||||| || |||||||| | || ||  ||| | |||
Db    1592 ATCGAAGCGGCTGGTGGCTCATTTCATTTGGCTAAATGTGAACTGGGGCAGAGCGCTGAT 1651

Qy    1443 GATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAATT 1502
           | |   || |||||||| |||||||| || || |||||||       ||||||| ||||
Db    1652 GCTGAATCGTACTCAGAACTTGAAGTTGGTGCGGATGATAAGAGAGTATTGGATCAAATC 1711

Qy    1503 ATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGAT---CACGATGCCGGGCAA 1559
           |||||||| || |||   |||||||||  | |||  |   ||     |       |  ||
```

```
Db   1712 ATTGATTCATTAACTCGGTTAGCTAATCCAAATGAAGATTATATATCCCCACATAGAGAA 1771

Qy   1560 GAAATTGAATTAGCTCTGAAGATAGGAAAAGTC--------AATGAGTATGAAACTGACG 1611
          | || | |  |  | ||||||||| || ||||||||        ||     || ||| || |
Db   1772 GCAAATAAGATCTCACTGAAGATTGGTAAAGTCCAGCAAGAAAATGAGATAAAAGAGAAG 1831

Qy   1612 TCACAATTGATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGC 1671
           |  || |||  ||  |    ||    ||||| |||||||| |||||| | || ||||| |
Db   1832 CCTGAAATGACGAA--AAAATCAGGTGTTTTGATTCTTGGTGCTGGACGTGTGTGTCGCC 1889

Qy   1672 CAGCTGCTGAGTTTCTGGCATCTTACCCAGACATATGTAC-------------------- 1711
          |||||||||| || || || ||      |  ||| |   |
Db   1890 CAGCTGCTGATTTCCTAGCTTCAGTTAGAACCATTTCGTCACAGCAATGGTACAAAACAT 1949

Qy   1712 -CTATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATCTTTGTATC 1770
            |  || |  ||     | ||   |    |  ||||||| || || || ||| ||||||
Db   1950 ATTTCGGAGCAGACTCTGAAGAGAAAACAGATGTTCATGTGATTGTCGCGTCTCTGTATC 2009

Qy   1771 AAAAAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCAGCTTGATG 1830
            || |||||  ||||||| ||||| ||||||  | ||  |   ||    | ||| ||||
Db   2010 TTAAGGATGCCAAAGAGACGGTTGAAGGTATTTCAGATGTAGAAGCAGTTCGGCTAGATG 2069

Qy   1831 TTGCTGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGC 1890
          |  |||||| || ||| ||     |   ||||||||||||||| ||||| | || ||
Db   2070 TATCTGATAGTGAAAGTCTCCTTAAGTATGTTTCTCAGGTTGATGTTGTCCTAAGTTTAT 2129

Qy   1891 TGCCTGCTAGTTTTCATGCTGCCATTGCAGGAGTATGCATAGAGTTGAAGAAGCACATGG 1950
          | ||||| |||| ||||||||   | |||     |||||| ||| |||||||||||  | |
Db   2130 TACCTGCAAGTTGTCATGCTGTTGTAGCAAAGACATGCATTGAGCTGAAGAAGCATCTCG 2189

Qy   1951 TAACGGCAAGCTATGTTGATGAATCCATGTCAAACTTGAGCCAAGCTGCCAAAGATGCAG 2010
          | || || ||||||||||||||    | || | |  ||     |    || ||   ||| |
Db   2190 TCACTGCTAGCTATGTTGATGATGAAACGTCCATGTTACATGAGAAGGCTAAGAGTGCTG 2249

Qy   2011 GTGTAACTATACTTTGTGAAATGGGCCTAGATCCTGGCATAGATCACTTGATGTCAATGA 2070
          |  |||| || ||  | |||||||| || || ||||| || |||||| ||||| | ||||
Db   2250 GGATAACGATTCTAGGCGAAATGGGACTGGACCCTGGAATCGATCACATGATGGCGATGA 2309

Qy   2071 AGATGATTGATGAAGCTCATGCACGAAAGGGAAAAATAAAGGCATTTACATCTTACTGTG 2130
          | |||||  | || ||||||     ||| || ||| | ||| | ||||| ||||| ||||
Db   2310 AAATGATCAACGATGCTCATATCAAAAAAGGGAAAGTGAAGTCTTTTACCTCTTATTGTG 2369

Qy   2131 GTGGATTGCCATCTCCAGCTGCAGCAAACAATCCGCTTGCCTATAAATTCAGTTGGAACC 2190
          | ||  | || ||||| |||||||||| ||||  | || |||||||| || |||||||
Db   2370 GAGGGCTTCCCTCTCCTGCTGCAGCAAATAATCCATTAGCATATAAATTTAGCTGGAACC 2429

Qy   2191 CAGCTGGTGCACTCCGGTCAGGGAAAAATCCTGCAGTCTACAAATTTCTTGGTGAGACGA 2250
          | ||||| ||| | ||  | ||  |||| || ||    ||||||      || || |  |
Db   2430 CTGCTGGAGCAATTCGAGCTGGTCAAAACCCCGCCAAATACAAAAGCAACGGCGACATAA 2489

Qy   2251 TCCATGTAGATGGTCATAACTTGTATGAATCAGCAAAGAGGCTCAGACTACGAGAGCTTC 2310
          | ||||| ||||||  | || | ||||| || ||    ||  || || ||   |   | ||||
Db   2490 TACATGTTGATGGGAAGAATCTCTATGATTCCGCGGCAAGATTCCGAGTACCTAATCTTC 2549

Qy   2311 CAGCTTTTGCTCTGGAACACTTGCCAAATCGGAATTCCTTGATATATGGTGACCTTTATG 2370
          ||||||||||  ||||     | ||||||||  | |||||| | || || || | |||||
Db   2550 CAGCTTTTGCATTGGAGTGTCTTCCAAATCGTGACTCCTTGGTTTACGGGGAACATTATG 2609

Qy   2371 GTATCTCCAAAGAAGCATCCACCATATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTG 2430
          | |||   |  |||||| | || |||| | | |  || :|  | || ||||| ||||||
Db   2610 GCATCGAGAGCGAAGCAACAACGATATTTCGTGGAACACTCAGATATGAAGGGTTTAGTA 2669

Qy   2431 AGATTATGGTAACCCTTTCCAAAACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGC 2490
           ||| |||| ||| ||||| |||  ||| ||||||||   || |    ||| | | ||
Db   2670 TGATAATGGCAACACTTTCGAAACTTGGATTCTTTGACAGTGAAGCAAATCAAGTACTCT 2729

Qy   2491 AAGATACTAGTCGTCCAACATATAAGGGTTTCCTTGATGAACTACTGAATAATATCTCCA 2550
              |   |   |    || | |   | | |  |   | |  | || |||||     |
Db   2730 CCACTGGAAAGAGGATTACGTTTGGTGCTCTTTTAAGTAACATTCTAAATAAGGATGCAG 2789

Qy   2551 CAATTAACACGGACTTAGATATTGAAGCTTCTGGTGGATACGATGATGACCTGATTGCCA 2610
            | | |  |               ||   |    |     || || ||  | |     |
Db   2790 ACAATGAATC---------------AGAGCCCCTAGCGGGAGAAGAAGAGATAAGCAAGA 2834

Qy   2611 GACTGTTGAAGCTCGGGTGTTGCAAAAATAAGGAAATAGCTGTTAAGACAGTCAAAACCA 2670
          || |  | ||||| ||   | |      ||||| |  ||  |  ||  | |||||| |
Db   2835 GAATTATCAAGCTTGGACATTCC------AAGGAGACTGCAGCCAAAGCTGCCAAAACAA 2888

Qy   2671 TCAAGTTCTTGGGACTACATGAAGAGACTCAAATACCTAAGGGTTGTTCGAGCCCATTTG 2730
          |    ||||||||  |  | |||||||    |  | ||       |||   |||  |||||
Db   2889 TTGTATTCTTGGGGTTCAACGAAGAGAGGGAGGTTCCATCACTGTGTAAAAGCGTATTTG 2948

Qy   2731 ATGTGATTTGCCAGCGAATGGAACAGAGGATGGCCTATGGCCACAATGAGCAAGACATGG 2790
          |||  | |||  | | |||||||| |||   | || |||  |   ||||| || |||||||
Db   2949 ATGCAACTTGTTACCTAATGGAAGAGAAACTAGCTTATTCCGGAAATGAACAGGACATGG 3008

Qy   2791 TACTGCTCCACCACGAAGTCGAGGTGGAATACCCGGACGGGCAACCCGCCGAAAAGCACC 2850
          | ||  | || |||||||| || ||||||| ||  ||  |  |||     || ||||||
Db   3009 TGCTTTTGCATCACGAAGTAGAAGTGGAATTCCTTGAAAGCAAACGTATAGAGAAGCACA 3068

Qy   2851 AAGCGACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCACCACTGCCATGGCGC 2910
            ||||| ||  |||| |||||| |  |  | |||||     | || || || |||||
Db   3069 CTGCGACTCTTTTGGAATTCGGGGACATCAAGAATGGACAAACAACAACCGCTATGGCCA 3128

Qy   2911 TGACCGTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAGGTCCAGA 2970
           ||| || || || || || || || || || ||||| ||  |  |  | ||| || |||
Db   3129 AGACTGTTGGGATCCCTGCAGCCATTGGAGCTCTGCTGTTAATTGAAGACAAGATCAAGA 3188

Qy   2971 CGAAAGGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAGCATTGGAGATCTTGG 3030
          | | ||||||  | ||||||||  || | | ||  | ||  | ||||| ||||| || |||
Db   3189 CAAGAGGAGTCTTAAGGCCTCTCGAAGCAGAGGTGTATTTGCCAGCTTTGGATATATTGC 3248

Qy   3031 AGTCGTCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGAAAGTTCCCTG 3081
          |  | |  || || |||||| | ||||| |  ||    | ||||  | |||
Db   3249 AAGCATATGGTATAAAGCTGATGGAGAAGGCAGAATGATCAAAGAACTCTG 3299
```

```
RESULT   5
BD010447
LOCUS       BD010447                3195 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Chimeric genes and methods for increasing the lysine content of the
             seeds  of plants Chimeric genes and methods for increasing the
            lysine content of the seeds  of plants.
ACCESSION   BD010447
VERSION     BD010447.1  GI:18638820
KEYWORDS    JP 2001502923-A/79.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 3195)
  AUTHORS   Falco,S.C., Iii,R.E.M. and Epelbaum,S.U.
  TITLE     Chimeric genes and methods for increasing the lysine content of the
            seeds of plants
  JOURNAL   Patent: JP 2001502923-A 79 06-MAR-2001;
            EI DU PONT DE NEMOURS AND CO
COMMENT     OS   Unidentified
            PN   JP 2001502923-A/79
            PD   06-MAR-2001
            PF   27-MAR-1998 JP 1998543284
            PR   27-MAR-1997 US     08/824627
            PI   SAVERIO CARL FALCO,RAYMOND ERVIN MCDEVITT III, PI    SABINE
            URSULA EPELBAUM
            PC   C12N9/06,C12N9/12,C12N9/88,C12P13/08,C12N15/82 CC
            Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .3195
            FT   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .3195
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      967 a    602 c    750 g    876 t
ORIGIN
```

```
Query Match             37.9%;  Score 1236.4;  DB 6;  Length 3195;
Best Local Similarity   64.6%;  Pred. No. 9.4e-308;
Matches 1999;  Conservative    1;  Mismatches 1027;  Indels   67;  Gaps    8;

Qy       15 CTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTGTGCAGCCAAGCACA 74
Db      121 CTTTTACACGGTGGGAAAGACAGAACCGGCATTTCCCGCATTGTGGTTCAGCCATCTGCT 180

Qy       75 AGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGATTTCAGAAGACCTG 134
Db      181 AAGCGTATCCATCATGATGCCTTGTATGAACATGTTGGGTGTGAAATTTCTGATGATTTG 240

Qy      135 TCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGATGATTCTTTCAGAT 194
Db      241 TCTGATTGTGGGCTTATACTTGGAATCAAACAACCTGAGCTAGAAATGATTCTTCCAGAG 300

Qy      195 AGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGAATATGCCACTGTTA 254
Db      301 AGAGCATACGCTTTCTTTTCACATACTCATAAGGCACAGAAAGAGAACATGCCTTTGTTG 360

Qy      255 GACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAATTGTTGGAGATGAT 314
Db      361 GATAAAATTCTTTCTGAGAGAGTGACTTTGTGTGATTATGAGCTCATTGTTGGGGATCAT 420

Qy      315 GGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGACTGATAGATTTCTTA 374
Db      421 GGGAAACGATTATTGGCGTTTGGTAAATATGCAGGCAGAGCTGGTCTTGTTGACTTCTTA 480

Qy      375 CATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCATTTCTCTCTCTGGGA 434
Db      481 CACGGACTTGGACAGCGATATCTAAGTCTAGGATACTCAACACCTTTCCTCTCGCTCGGT 540

Qy      435 CAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCATTGTCGTTGCAGAA 494
Db      541 GCATCGTATATGTATTCCTCATTGGCTGCTGCAAAAGCCGCTGTAATTTCTGTTGGTGAA 600

Qy      495 GAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATAGTGTTTGTGTTCACTGGA 554
Db      601 GAAATTGCAAGCCAGGGACTGCCATTAGGAATCTGCCCTCTTGTATTTGTCTTCACCGGA 660

Qy      555 GTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGCCCCATACCTTTGTT 614
Db      661 ACAGGAAATGTTTCTCTGGGGGCGCAAGAAATTTTCAAGCTTCTTCCTCACACTTTTGTT 720

Qy      615 GATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTAAGCAATCTCAG--- 671
Db      721 GAACCAAGCAAACTTCCTGAACTATTTGTAAAAGACAAAGGAATTAGTCAAAATGGGATT 780

Qy      672 TCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAGACATAGTTTCT 731
Db      781 TCAACAAAGCGAGTCTATCAAGTATATGGTTGTATTATTACCAGCCAAGACATGGTTGAA 840

Qy      732 CACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATCCAGAACACTAC 791
Db      841 CACAAAGATCCATCAAAGTCATTCGACAAAGCCGACTATTATGCACACCCGGAACATTAC 900

Qy      792 ACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAAACTGTATGTAT 851
Db      901 AATCCAGTTTTCCACGAAAAGATATCGCCATATACGTCTGTTCTTGTAAACTGTATGTAC 960

Qy      852 TGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAATTGATGGAGACT 911
Db      961 TGGGAGAAGAGGTTTCCCTGTCTTCTGAGCACAAAACAGCTTCAAGATTTAACAAAAAAA 1020

Qy      912 GGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTTCCATTGAATTT 971
Db     1021 GGACTCCCACTAGTAGGCATATGTGATATAACTTGTGACATCGGTGGCTCCATTGAATTT 1080

Qy      972 ATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTTCTAAGAATTCA 1031
Db     1081 GTTAACCGAGCTACTTTAATCGATTCCCCTTTCTTCAGGTTTAATCCCTCGAACAATTCA 1140

Qy     1032 TACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACATTCTCCCTACA 1091
Db     1141 TACTACGATGACATGGATGGGGATGGCGTACTATGCATGGCTGTTGACATTTTACCCACA 1200

Qy     1092 GAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGACTTGTTGCTAGT 1151
Db     1201 GAATTTGCAAAAGAGGCATCCCAGCATTTTGGAGATATTCTTTCCGGATTTGTCGGTAGT 1260

Qy     1152 TTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAGCTTGCATTGCA 1211
Db     1261 TTGGCTTCAATGACTGAAATTTCAGATCTACCAGCACATCTGAAGAGGGCTTGCATAAGC 1320

Qy     1212 CATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAAATACTATGATA 1271
Db     1321 TATAGGGGAGAATTGACATCTTTGTATGAGTATATTCCACGTATGAGGAAGTCAAATCCA 1380

Qy     1272 GATTTGGCACCCGCAAAAACAAATCCATTGCCTG---------ACAAGAAGTATAGCACC 1322
Db     1381 GAAGAGGCACAAGATAATATTATCGCCAACGGGGTTTCCAGCCAGAGAACATTCAACATA 1440

Qy     1323 CTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACATC 1382
Db     1441 TTGGTATCTCTGAGCGGACACCTATTTGATAAGTTTCTGATAAACGAAGCTCTTGATATG 1500

Qy     1383 ATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGGAT 1442
Db     1501 ATCGAAGCGGCTGGTGGCTCATTTCATTTGGCTAAATGTGAACTGGGGCAGAGCGCTGAT 1560

Qy     1443 GATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAATT 1502
Db     1561 GCTGAATCGTACTCAGAACTTGAAGTTGGTGCGGATGATAAGAGAGTATTGGATCAAATC 1620

Qy     1503 ATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGC---CGGGCAA 1559
Db     1621 ATTGATTCATTAACTCGGTTAGCTAATCCAAATGAAGATTATATATCCCCACATAGAGAA 1680

Qy     1560 GAAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAG--------TATGAAACTGACG 1611
Db     1681 GCAAATAAGATCTCACTGAAGATTGGTAAAGTCCAGCAAGAAAATGAGATAAAAGAGAAG 1740

Qy     1612 TCACAATTGATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGC 1671
Db     1741 CCTGAAATGACGAA--AAAATCAGGTGTTTTGATTCTTGGTGCTGGACGTGTGTGTCGCC 1798

Qy     1672 CAGCTGCTGAGTTTCTGGCATCTTACCCAGACATATGTAC-------------------- 1711
Db     1799 CAGCTGCTGATTTCCTAGCTTCAGTTAGAACCATTTCGTCACAGCAATGGTACAAAACAT 1858

Qy     1712 -CTATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATCTTTGTATC 1770
Db     1859 ATTTCGGAGCAGACTCTGAAGAGAAAACAGATGTTCATGTGATTGTCGCGTCTCTGTATC 1918

Qy     1771 AAAAAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCAGCTTGATG 1830
Db     1919 TTAAGGATGCCAAAGAGACGGTTGAAGGTATTTCAGATGTAGAAGCAGTTCGGCTAGATG 1978

Qy     1831 TTGCTGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGC 1890
Db     1979 TATCTGATAGTGAAAGTCTCCTTAAGTATGTTTCTCAGGTTGATGTTGTCCTAAGTTTAT 2038

Qy     1891 TGCCTGCTAGTTTTCATGCTGCCATTGCAGGAGTATGCATAGAGTTGAAGAAGCACATGG 1950
Db     2039 TACCTGCAAGTTGTCATGCTGTTGTAGCAAAGACATGCATTGAGCTGAAGAAGCATCTCG 2098

Qy     1951 TAACGGCAAGCTATGTTGATGAATCCATGTCAAACTTGAGCCAAGCTGCCAAAGATGCAG 2010
Db     2099 TCACTGCTAGCTATGTTGATGATGAAACGTCCATGTTACATGAGAAGGCTAAGAGTGCTG 2158

Qy     2011 GTGTAACTATACTTTGTGAAATGGGCCTAGATCCTGGCATAGATCACTTGATGTCAATGA 2070
Db     2159 GGATAACGATTCTAGGCGAAATGGGACTGGACCCTGGAATCGATCACATGATGGCGATGA 2218

Qy     2071 AGATGATTGATGAAGCTCATGCACGAAAGGGAAAAATAAAGGCATTTACATCTTACTGTG 2130
```

```
Db     2219 AAATGATCAACGATGCTCATATCAAAAAAGGGAAAGTGAAGTCTTTTACCTCTTATTGTG 2278

Qy     2131 GTGGATTGCCATCTCCAGCTGCAGCAAACAATCCGCTTGCCTATAAATTCAGTTGGAACC 2190
            | ||  | || ||||| ||||||||||| |||||  | || |||||||| || |||||||
Db     2279 GAGGGCTTCCCTCTCCTGCTGCAGCAAATAATCCATTAGCATATAAATTTAGCTGGAACC 2338

Qy     2191 CAGCTGGTGCACTCCGGTCAGGGAAAAATCCTGCAGTCTACAAATTTCTTGGTGAGACGA 2250
            | ||||| ||| | ||  | ||  |||| || ||    ||||||      || || |  |
Db     2339 CTGCTGGAGCAATTCGAGCTGGTCAAAACCCCGCCAAATACAAAAGCAACGGCGACATAA 2398

Qy     2251 TCCATGTAGATGGTCATAACTTGTATGAATCAGCAAAGAGGCTCAGACTACGAGAGCTTC 2310
            | ||||| |||||  | ||  | ||||| || ||    ||  || || |||   | ||||
Db     2399 TACATGTTGATGGGAAGAATCTCTATGATTCCGCGGCAAGATTCCGAGTACCTAATCTTC 2458

Qy     2311 CAGCTTTTGCTCTGGAACACTTGCCAAATCGGAATTCCTTGATATATGGTGACCTTTATG 2370
            ||||||||||  ||||    || ||||||||  | |||||| | || || || | |||||
Db     2459 CAGCTTTTGCATTGGAGTGTTTTCCAAATCGTGACTCCTTGGTTTACGGGGAACATTATG 2518

Qy     2371 GTATCTCCAAAGAAGCATCCACCATATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTG 2430
            | |||   |  |||||| | || |||| | | |  || :|  | || ||||| ||||||
Db     2519 GCATCGAGAGCGAAGCAACAACGATATTTCGTGGAACACTCAGATATGAAGGGTTTAGTA 2578

Qy     2431 AGATTATGGTAACCCTTTCCAAAACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGC 2490
             ||| |||| ||| ||||| |||  ||| ||||||||   || |    ||| | | ||
Db     2579 TGATAATGGCAACACTTTCGAAACTTGGATTCTTTGACAGTGAAGCAAATCAAGTACTCT 2638

Qy     2491 AAGATACTAGTCGTCCAACATATAAGGGTTTCCTTGATGAACTACTGAATAATATCTCCA 2550
                |   |   |    || | |   | | |  |   | |  | || |||||     |
Db     2639 CCACTGGAAAGAGGATTACGTTTGGTGCTCTTTTAAGTAACATTCTAAATAAGGATGCAG 2698

Qy     2551 CAATTAACACGGACTTAGATATTGAAGCTTCTGGTGGATACGATGATGACCTGATTGCCA 2610
              | | |  |               ||   |    |     || || ||  | |     |
Db     2699 ACAATGAATC---------------AGAGCCCCTAGCGGGAGAAGAAGAGATAAGCAAGA 2743

Qy     2611 GACTGTTGAAGCTCGGGTGTTGCAAAAATAAGGAAATAGCTGTTAAGACAGTCAAAACCA 2670
            || |  | ||||| ||   || |      ||||| |  || |  ||  | | |||||| |
Db     2744 GAATTATCAAGCTTGGACATTCC------AAGGAGACTGCAGCCAAAGCTGCCAAAACAA 2797

Qy     2671 TCAAGTTCTTGGGACTACATGAAGAGACTCAAATACCTAAGGGTTGTTCGAGCCCATTTG 2730
            |    ||||||||  |  | |||||||   |  | ||       |||   |||  |||||
Db     2798 TTGTATTCTTGGGGTTCAACGAAGAGAGGGAGGTTCCATCACTGTGTAAAAGCGTATTTG 2857

Qy     2731 ATGTGATTTGCCAGCGAATGGAACAGAGGATGGCCTATGGCCACAATGAGCAAGACATGG 2790
            |||  | |||  | | ||||||| |||   | || |||  |   ||||| || |||||||
Db     2858 ATGCAACTTGTTACCTAATGGAAGAGAAACTAGCTTATTCCGGAAATGAACAGGACATGG 2917

Qy     2791 TACTGCTCCACCACGAAGTCGAGGTGGAATACCCGGACGGGCAACCCGCCGAAAAGCACC 2850
            | ||  | || |||||||| || ||||||| ||  ||  |  |||     || ||||||
Db     2918 TGCTTTTGCATCACGAAGTAGAAGTGGAATTCCTTGAAAGCAAACGTATAGAGAAGCACA 2977

Qy     2851 AAGCGACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCACCACTGCCATGGCGC 2910
              ||||| ||  |||| |||||| |  |  | |||||     | || || || |||||
Db     2978 CTGCGACTCTTTTGGAATTCGGGGACATCAAGAATGGACAAACAACAACCGCTATGGCCA 3037

Qy     2911 TGACCGTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAGGTCCAGA 2970
             ||| || || || || ||||| || || || ||| | ||  |  |  | ||| || |||
Db     3038 AGACTGTTGGGATCCCTGCAGCCATTGGAGCTCTGGTGTTAATTGAAGACAAGATCAAGA 3097

Qy     2971 CGAAAGGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAGCATTGGAGATCTTGG 3030
            | | ||||||  | ||||||||  || | ||  | ||  | ||||| ||||| |  |||
Db     3098 CAAGAGGAGTCTTAAGGCCTCTCGAAGCAGAGGTGTATTTGCCAGCTTTGGATATATTGC 3157

Qy     3031 AGTCGTCGGGCATCAAGCTGGTTGAGAAAGTGGA 3064
            |  | |  || || |||||| | ||||| |  ||
Db     3158 AAGCATATGGTATAAAGCTGATGGAGAAGGCAGA 3191


RESULT    6
ATU90522
LOCUS       ATU90522                3267 bp    mRNA    linear   PLN 27-SEP-2001
DEFINITION  Arabidopsis thaliana lysine-ketoglutarate reductase/saccharopine
            dehydrogenase (LKR/SDH) mRNA, complete cds.
ACCESSION   U90522
```

```
VERSION     U90522.1  GI:2076883
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 3267)
  AUTHORS   Tang,G., Miron,D., Zhu-Shimoni,J.X. and Galili,G.
  TITLE     Regulation of lysine catabolism through lysine-ketoglutarate
            reductase and saccharopine dehydrogenase in Arabidopsis
  JOURNAL   Plant Cell 9 (8), 1305-1316 (1997)
  MEDLINE   97432144
   PUBMED   9286108
REFERENCE   2  (bases 1 to 3267)
  AUTHORS   Tang,G., Miron,D., Zhu-Shimoni,J.X. and Galili,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1997) Dept. of Plant Genetics, The Weizmann
            Institute of Science, Rehovot 76100, Israel
FEATURES             Location/Qualifiers
     source          1..3267
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /tissue_type="seedling hypocotyls"
     gene            1..3267
                     /gene="LKR/SDH"
     CDS             63..3257
                     /gene="LKR/SDH"
                     /EC_number="1.5.1.8"
                     /EC_number="1.5.1.9"
                     /function="catalyses the first two steps in lysine
                     catabolism"
                     /note="bifuctional enzyme; aminoadipic semialdehyde
                     synthase"
                     /codon_start=1
                     /product="lysine-ketoglutarate reductase/saccharopine
                     dehydrogenase"
                     /protein_id="AAB53975.1"
                     /db_xref="GI:2076884"
                     /translation="MNSNGHEEEKKLGNGVVGILAETVNKWERRTPLTPSHCARLLHG
                     GKDRTGISRIVVQPSAKRIHHDALYEDVGCEISDDLSDCGLILGIKQPELEMILPERA
                     YAFFSHTHKAQKENMPLLDKILSERVTLCDYELIVGDHGKRLLAFGKYAGRAGLVDFL
                     HGLGQRYLSLGYSTPFLSLGASYMYSSLAAAKAAVISVGEEIASQGLPLGICPLVFVF
                     TGTGNVSLGAQEIFKLLPHTFVEPSKLPELFVKDKGISQNGISTKRVYQVYGCIITSQ
                     DMVEHKDPSKSFDKADYYAHPEHYNPVFHEKISPYTSVLVNCMYWEKSFPCLLSTKQL
                     QDLTKKGLPLVGICDITCDIGGSIEFVNRATLIDSPFFRFNPSNNSYYDDMDGDGVLC
                     MAVDILPTEFAKEASQHFGDILSGFVGSLASMTEISDLPAHLKRACISYRGELTSLYE
                     YIPRMRKSNPEEAQDNIIANGVSSQRTFNILVSLSGHLFDKFLINEALDMIEAAGGSF
                     HLAKCELGQSADAESYSELEVGADDKRVLDQIIDSLTRLANPNEDYISPHREANKISL
                     KIGKVQQENEIKEKPEMTKKSGVLILGAGRVCRPAADFLASVRTISSQQWYKTYFGAD
                     SEEKTDVHVIVASLYLKDAKETVEGISDVEAVRLDVSDSESLLKYVSQVDVVLSLLPA
                     SCHAVVAKTCIELKKHLVTASYVDDETSMLHENAKSAGITILGEMGLDPGIDPMMAKK
                     MINDAPITKGKVKSFTSYCGGLPSRAAANNPLAYKFSWNPAGAIRAGQNPAKYKSNGD
                     IIHVDGKNLYDSAARFRVPNLPAFALECFPNRDSLVYGEHYGIESEATTIFRGTLRYE
                     GFSMIMATLSKLGFFDSEANQVLSTGKRITFGALLSNILNKDADNESEPLAGEEEISK
                     RIIKLGHSKETAAKAAKTIVFLGFNEEREVPSLCKSVFDATCYLMEEKLAYSGNEQDM
                     VLLHHEVEVEFLESKRIEKHTATLLEFGDIKNGQTTTAMAKTVGIPAAIGALLLIEDK
                     IKTRGVLRPLEAEVYLPALDILQAYGIKLMEKAE"
BASE COUNT      993 a    620 c    761 g    893 t
ORIGIN


  Query Match             37.7%;  Score 1232;  DB 8;  Length 3267;
  Best Local Similarity   64.4%;  Pred. No. 1.3e-306;
  Matches 2000;  Conservative    1;  Mismatches 1036;  Indels   67;  Gaps    8;

Qy       15 CTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTGTGCAGCCAAGCACA 74
            ||  ||   || || || ||  || ||  |  |  |   ||| ||| | || ||||||    |
Db      183 CTTTTACACGGTGGGAAAGACAGAACCGGCATTTCCCGCATTGTGGTTCAGCCATCTGCT 242

Qy       75 AGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGATTTCAGAAGACCTG 134
            | | | ||||||||||| ||   |||||| ||||  || || || ||||| || ||  ||
```

```
Db      243 AAGCGTATCCATCATGATGCCTTGTATGAAGATGTTGGGTGTGAAATTTCTGATGATTTG 302

Qy      135 TCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGATGATTCTTTCAGAT 194
Db      303 TCTGATTGTGGGCTTATACTTGGAATCAAACAACCTGAGCTAGAAATGATTCTTCCAGAG 362

Qy      195 AGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGAATATGCCACTGTTA 254
Db      363 AGAGCATACGCTTTCTTTTCACATACTCATAAGGCACAGAAAGAGAACATGCCTTTGTTG 422

Qy      255 GACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAATTGTTGGAGATGAT 314
Db      423 GATAAAATTCTTTCTGAGAGAGTGACTTTGTGTGATTATGAGCTCATTGTTGGGGATCAT 482

Qy      315 GGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGACTGATAGATTTCTTA 374
Db      483 GGGAAACGATTATTGGCGTTTGGTAAATATGCAGGCAGAGCTGGTCTTGTTGACTTCTTA 542

Qy      375 CATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCATTTCTCTCTCTGGGA 434
Db      543 CACGGACTTGGACAGCGATATCTAAGTCTAGGATACTCAACACCTTTCCTCTCGCTCGGT 602

Qy      435 CAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCATTGTCGTTGCAGAA 494
Db      603 GCATCGTATATGTATTCCTCATTGGCTGCTGCAAAAGCCGCTGTAATTTCTGTTGGTGAA 662

Qy      495 GAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATAGTGTTTGTGTTCACTGGA 554
Db      663 GAAATTGCAAGCCAGGGACTGCCATTAGGAATCTGCCCTCTTGTATTTGTCTTCACCGGA 722

Qy      555 GTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGCCCCATACCTTTGTT 614
Db      723 ACAGGAAATGTTTCTCTGGGGGCGCAAGAAATTTTCAAGCTTCTTCCTCACACTTTTGTT 782

Qy      615 GATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTAAGCAATCTCAG--- 671
Db      783 GAACCAAGCAAACTTCCTGAACTATTTGTAAAAGACAAAGGAATTAGTCAAAATGGGATT 842

Qy      672 TCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAGACATAGTTTCT 731
Db      843 TCAACAAAGCGAGTCTATCAAGTATATGGTTGTATTATTACCAGCCAAGACATGGTTGAA 902

Qy      732 CACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATCCAGAACACTAC 791
Db      903 CACAAAGATCCATCAAAGTCATTCGACAAAGCCGACTATTATGCACACCCGGAACATTAC 962

Qy      792 ACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAAACTGTATGTAT 851
Db      963 AATCCAGTTTTCCACGAAAAGATATCGCCATATACGTCTGTTCTTGTAAACTGTATGTAC 1022

Qy      852 TGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAATTGATGGAGACT 911
Db     1023 TGGGAGAAGAGTTTTCCCTGTCTTCTGAGCACAAAACAGCTTCAAGATTTAACAAAAAAA 1082

Qy      912 GGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTTCCATTGAATTT 971
Db     1083 GGACTCCCACTAGTAGGCATATGTGATATAACTTGTGACATCGGTGGCTCCATTGAATTT 1142

Qy      972 ATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTTCTAAGAATTCA 1031
Db     1143 GTTAACCGAGCTACTTTAATCGATTCCCCTTTCTTCAGGTTTAATCCCTCGAACAATTCA 1202

Qy     1032 TACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACATTCTCCCTACA 1091
Db     1203 TACTACGATGACATGGATGGGGATGGCGTACTATGCATGGCTGTTGACATTTTACCCACA 1262

Qy     1092 GAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGACTTGTTGCTAGT 1151
Db     1263 GAATTTGCAAAAGAGGCATCCCAGCATTTTGGAGATATTCTTTCCGGATTTGTCGGTAGT 1322

Qy     1152 TTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAGCTTGCATTGCA 1211
Db     1323 CTGGCTTCAATGACTGAAATTTCAGATCTACCAGCACATCTGAAGAGGGCTTGCATAAGC 1382
```

```
Qy     1212 CATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAAATACTATGATA 1271
Db     1383 TATAGGGGAGAATTGACATCTTTGTATGAGTATATTCCACGTATGAGGAAGTCAAATCCA 1442

Qy     1272 GATTTGGCACCCGCAAAAACAAATCCATTGCCTG---------ACAAGAAGTATAGCACC 1322
Db     1443 GAAGAGGCACAAGATAATATTATCGCCAACGGGGTTTCCAGCCAGAGAACATTCAACATA 1502

Qy     1323 CTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACATC 1382
Db     1503 TTGGTATCTCTGAGCGGACACCTATTTGATAAGTTTCTGATAAACGAAGCTCTTGATATG 1562

Qy     1383 ATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGGAT 1442
Db     1563 ATCGAAGCGGCTGGTGGCTCATTTCATTTGGCTAAATGTGAACTGGGGCAGAGCGCTGAT 1622

Qy     1443 GATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAATT 1502
Db     1623 GCTGAATCGTACTCAGAACTTGAAGTTGGTGCGGATGATAAGAGAGTATTGGATCAAATC 1682

Qy     1503 ATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCC---GGGCAA 1559
Db     1683 ATTGATTCATTAACTCGGTTAGCTAATCCAAATGAAGATTATATATCCCCACATAGAGAA 1742

Qy     1560 GAAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGT--------ATGAAACTGACG 1611
Db     1743 GCAAATAAGATCTCACTGAAGATTGGTAAAGTCCAGCAAGAAAATGAGATAAAAGAGAAG 1802

Qy     1612 TCACAATTGATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGC 1671
Db     1803 CCTGAAATGACGAAAAAA--TCAGGTGTTTTGATTCTTGGTGCTGGACGTGTGTGTCGCC 1860

Qy     1672 CAGCTGCTGAGTTTCTGGCATCTTACCCAGACATATGTAC-------------------- 1711
Db     1861 CAGCTGCTGATTTCCTAGCTTCAGTTAGAACCATTTCGTCACAGCAATGGTACAAAACAT 1920

Qy     1712 -CTATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATCTTTGTATC 1770
Db     1921 ATTTCGGAGCAGACTCTGAAGAGAAAACAGATGTTCATGTGATTGTCGCGTCTCTGTATC 1980

Qy     1771 AAAAAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCAGCTTGATG 1830
Db     1981 TTAAGGATGCCAAAGAGACGGTTGAAGGTATTTCAGATGTAGAAGCAGTTCGGCTAGATG 2040

Qy     1831 TTGCTGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGC 1890
Db     2041 TATCTGATAGTGAAAGTCTCCTTAAGTATGTTTCTCAGGTTGATGTTGTCCTAAGTTTAT 2100

Qy     1891 TGCCTGCTAGTTTTCATGCTGCCATTGCAGGAGTATGCATAGAGTTGAAGAAGCACATGG 1950
Db     2101 TACCTGCAAGTTGTCATGCTGTTGTAGCAAAGACATGCATTGAGCTGAAGAAGCATCTCG 2160

Qy     1951 TAACGGCAAGCTATGTTGATGAATCCATGTCAAACTTGAGCCAAGCTGCCAAAGATGCAG 2010
Db     2161 TCACTGCTAGCTATGTTGATGATGAAACGTCCATGTTACATGAGAACGCTAAGAGTGCTG 2220

Qy     2011 GTGTAACTATACTTTGTGAAATGGGCCTAGATCCTGGCATAGATCACTTGATGTCAATGA 2070
Db     2221 GGATAACGATTCTAGGCGAAATGGGACTGGACCCTGGAATCGATCCCATGATGGCGAAGA 2280

Qy     2071 AGATGATTGATGAAGCTCATGCACGAAAGGGAAAAATAAAGGCATTTACATCTTACTGTG 2130
Db     2281 AAATGATCAACGATGCTCCTATCACAAAAGGGAAAGTGAAGTCTTTTACCTCTTATTGTG 2340

Qy     2131 GTGGATTGCCATCTCCAGCTGCAGCAAACAATCCGCTTGCCTATAAATTCAGTTGGAACC 2190
Db     2341 GAGGGCTTCCCTCTCGTGCTGCAGCAAATAATCCATTAGCATATAAATTTAGCTGGAACC 2400

Qy     2191 CAGCTGGTGCACTCCGGTCAGGGAAAAATCCTGCAGTCTACAAATTTCTTGGTGAGACGA 2250
Db     2401 CTGCTGGAGCAATTCGAGCTGGTCAAAACCCCGCCAAATACAAAAGCAACGGCGACATAA 2460
```

```
Qy     2251 TCCATGTAGATGGTCATAACTTGTATGAATCAGCAAAGAGGCTCAGACTACGAGAGCTTC 2310
            | ||||| |||||  | ||  | ||||| || ||    ||  || || |||   | ||||
Db     2461 TACATGTTGATGGGAAGAATCTCTATGATTCCGCGGCAAGATTCCGAGTACCTAATCTTC 2520

Qy     2311 CAGCTTTTGCTCTGGAACACTTGCCAAATCGGAATTCCTTGATATATGGTGACCTTTATG 2370
            ||||||||||  ||||    || ||||||||  | |||||| | || || || | |||||
Db     2521 CAGCTTTTGCATTGGAGTGTTTTCCAAATCGTGACTCCTTGGTTTACGGGGAACATTATG 2580

Qy     2371 GTATCTCCAAAGAAGCATCCACCATATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTG 2430
            | |||   |  |||||| | || |||| | | |  || :|  | || ||||| ||||||
Db     2581 GCATCGAGAGCGAAGCAACAACGATATTTCGTGGAACACTCAGATATGAAGGGTTTAGTA 2640

Qy     2431 AGATTATGGTAACCCTTTCCAAAACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGC 2490
             ||| |||| ||| ||||| |||  ||| ||||||||   || |    ||| | | ||
Db     2641 TGATAATGGCAACACTTTCGAAACTTGGATTCTTTGACAGTGAAGCAAATCAAGTACTCT 2700

Qy     2491 AAGATACTAGTCGTCCAACATATAAGGGTTTCCTTGATGAACTACTGAATAATATCTCCA 2550
                |   |   |    || | |   | | |  |   | |  | || |||||     |
Db     2701 CCACTGGAAAGAGGATTACGTTTGGTGCTCTTTTAAGTAACATTCTAAATAAGGATGCAG 2760

Qy     2551 CAATTAACACGGACTTAGATATTGAAGCTTCTGGTGGATACGATGATGACCTGATTGCCA 2610
              | | |  |               ||   |    |     || || ||  | |     |
Db     2761 ACAATGAATC---------------AGAGCCCCTAGCGGGAGAAGAAGAGATAAGCAAGA 2805

Qy     2611 GACTGTTGAAGCTCGGGTGTTGCAAAAATAAGGAAATAGCTGTTAAGACAGTCAAAACCA 2670
            || |  | ||||| ||   || |      ||||| |  || |  ||  |  | |||||| |
Db     2806 GAATTATCAAGCTTGGACATTCC------AAGGAGACTGCAGCCAAAGCTGCCAAAACAA 2859

Qy     2671 TCAAGTTCTTGGGACTACATGAAGAGACTCAAATACCTAAGGGTTGTTCGAGCCCATTTG 2730
            |    ||||||||  |  | |||||||   |  | ||       |||   |||  |||||
Db     2860 TTGTATTCTTGGGGTTCAACGAAGAGAGGGAGGTTCCATCACTGTGTAAAAGCGTATTTG 2919

Qy     2731 ATGTGATTTGCCAGCGAATGGAACAGAGGATGGCCTATGGCCACAATGAGCAAGACATGG 2790
            |||  | |||  | | ||||||| |||   | || |||  |   ||||| || |||||||
Db     2920 ATGCAACTTGTTACCTAATGGAAGAGAAACTAGCTTATTCCGGAAATGAACAGGACATGG 2979

Qy     2791 TACTGCTCCACCACGAAGTCGAGGTGGAATACCCGGACGGCAACCCGCCGAAAAGCACC 2850
            | ||  | || |||||||| || ||||||| ||  ||  |  |||     || ||||||
Db     2980 TGCTTTTGCATCACGAAGTAGAAGTGGAATTCCTTGAAAGCAAACGTATAGAGAAGCACA 3039

Qy     2851 AAGCGACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCACCACTGCCATGGCGC 2910
              ||||| ||  |||| |||||| |  |  | |||||     | || || || |||||
Db     3040 CTGCGACTCTTTTGGAATTCGGGGACATCAAGAATGGACAAACAACAACCGCTATGGCCA 3099

Qy     2911 TGACCGTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAGGTCCAGA 2970
             ||| || || || || ||||| || || || ||||| ||  |  |  | ||| || |||
Db     3100 AGACTGTTGGGATCCCTGCAGCCATTGGAGCTCTGCTGTTAATTGAAGACAAGATCAAGA 3159

Qy     2971 CGAAAGGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAGCATTGGAGATCTTGG 3030
            | | ||||||  | ||||||||  || | ||  | ||  | ||||| ||||| || |||
Db     3160 CAAGAGGAGTCTTAAGGCCTCTCGAAGCAGAGGTGTATTTGCCAGCTTTGGATATATTGC 3219

Qy     3031 AGTCGTCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGAAAG 3074
            |  | |  || || |||||| | ||||| |  ||    | ||||
Db     3220 AAGCATATGGTATAAAGCTGATGGAGAAGGCAGAATGATCAAAG 3263


RESULT   7
BD010452
LOCUS       BD010452                2582 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Chimeric genes and methods for increasing the lysine content of the
             seeds  of plants Chimeric genes and methods for increasing the
            lysine content of the seeds  of plants.
ACCESSION   BD010452
VERSION     BD010452.1  GI:18638825
KEYWORDS    JP 2001502923-A/84.
SOURCE      Glycine max.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 2582)
```

```
  AUTHORS   Falco,S.C., Iii,R.E.M. and Epelbaum,S.U.
  TITLE     Chimeric genes and methods for increasing the lysine content of the
            seeds of plants
  JOURNAL   Patent: JP 2001502923-A 84 06-MAR-2001;
            EI DU PONT DE NEMOURS AND CO
COMMENT     OS   Glycine max (soybean)
            PN   JP 2001502923-A/84
            PD   06-MAR-2001
            PF   27-MAR-1998 JP 1998543284
            PR   27-MAR-1997 US    08/824627
            PI-  SAVERIO CARL FALCO,RAYMOND ERVIN MCDEVITT III, PI   SABINE
            URSULA EPELBAUM
            PC   C12N9/06,C12N9/12,C12N9/88,C12P13/08,C12N15/82 CC
            Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   CDS             3. .2357.
FEATURES             Location/Qualifiers
     source          1. .2582
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
BASE COUNT      798 a    482 c    562 g    739 t      1 others
ORIGIN


  Query Match             27.8%;  Score 907;  DB 6;  Length 2582;
  Best Local Similarity   62.7%;  Pred. No. 8.7e-223;
  Matches 1531;  Conservative    2;  Mismatches  857;  Indels   51;  Gaps    6;

Qy      727 TTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATCCAGAAC 786
            ||   | ||| |||| |   | |   ||||||||||  ||||| ||| | || || || |
Db        1 TTGAACCCAAAGATCACGTGATAGTGTTTGACAAAGCAGACTACTATTCACACCCTGAGC 60

Qy      787 ACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAAACTGTA 846
            | ||||  ||   ||| ||||||| ||| || |||||||||||||| || || || || |
Db       61 ATTACAATCCCACTTTCCATGAAAAAATAGCACCATATGCATCTGTTATTGTCAATTGCA 120

Qy      847 TGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAATTGATGG 906
            ||||||||||||| || ||||| | ||| |  |      | ||| | ||  | || ||||
Db      121 TGTATTGGGAGAAAAGATTTCCTCAATTGCCGAGCTATAAGCAGATGCAAGACTTAATGG 180

Qy      907 AGACTGGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTTCCATTG 966
                 ||  | ||  | || ||  |   ||||||||| |||||||| || ||||| ||||
Db      181 GCCGGGGGAGCCCCCTTGTTGGAATAGCTGACATAACGTGTGATATAGGGGGTTCAATTG 240

Qy      967 AATTTATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTTCTAAGA 1026
            | ||| | |||    |||| ||||| ||    || |||||  | |||||||| |  |  |
Db      241 AGTTTGTTAACCGCGGTACTTCAATTGATTCACCCTTCTTCAGATATGATCCCTTAACAA 300

Qy     1027 ATTCATACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACATTCTCC 1086
            |||| ||||||||||||||||| ||    |||||| | ||||| |||||||||||||| |
Db      301 ATTCCTACCATGATGATATGGAGGGGAATGGAGTGATATGCTTAGCTGTTGACATTCTTC 360

Qy     1087 CTACAGAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGACTTGTTG 1146
            | ||||||||  | || ||||| ||||||||||||||||||||||| ||   | ||||||
Db      361 CAACAGAATTTGCAAAGGAGGCTTCCCAACATTTTGGAAACATACTTTCCCAATTTGTTG 420

Qy     1147 CTAGTTTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAGCTTGCA 1206
              | |||||| || |  |   |     || |  | ||| |  |||| || ||||||||||
Db      421 TAAATTTGGCTTCTGCTACAGACATTACAAAGTTGCCTGCTCACTTAAGGAGAGCTTGCA 480

Qy     1207 TTGCACATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAA------ 1260
            | || |||   ||     ||||  | || ||||| ||||||||  | ||| | |
Db      481 TAGCCCATAAAGGAGTGCTAACCTCCTTATATGATTATATCCCACGCATGCGGAGTTCTG 540

Qy     1261 ATACTATGATAGATTTGGCACCCGCAAAAACAAATCCATTGCCTGACAAGAAGTATAGCA 1320
            || |   |  ||  |  | |  |||| |||    || ||      | | |||||| |  |
Db      541 ATTCAGAGGAAGTATCAGAAAACGCAGAAAATTCTCTATCCAACAAAAGGAAGTACAATA 600

Qy     1321 CCCTGGTATCTCTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGCTTTGGACA 1380
                ||| ||||| ||||| ||| |||||||| |||| || |||||||| || || || |
Db      601 TATCGGTGTCTCTGAGTGGTCACTTATTTGATCAGTTTCTGATAAATGAGGCCTTAGATA 660
```

```
Qy    1381 TCATTGAGACAGCTGGAGGTTCATTTCACTTGGTTAGATGTGAAGTTGGACAAAGCACGG 1440
           | |||||  | || ||||| || || |||||| || |  ||  | || || || ||||  |
Db     661 TTATTGAAGCTGCAGGAGGCTCCTTCCACTTAGTCAACTGCCATGTGGGTCAGAGCATTG 720

Qy    1441 ATGATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAA 1500
           | |   | |||| ||| || |||||||| || |||||| | |  ||     ||||| |||
Db     721 AAGCCGTATCATTCTCTGAACTTGAAGTTGGTGCAGATAACAGGGCTGTTCTGGATCAAA 780

Qy    1501 TTATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCAAG 1560
           | |||||||| || ||| || | |||| |  |  ||     |||     | |    ||||
Db     781 TCATTGATTCTTTAACTGCTATTGCTAGTCCAACTGAACATGATAGATTTTCAAATCAAG 840

Qy    1561 A---AATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAA 1617
           |   || | || |  | || ||| | || |||||  | || |||  |  ||      | 
Db     841 ATTCAAGTAAAATTTCACTTAAGCTTGGTAAAGTTGAAGAGAATGGCATAGAGAAGGAAT 900

Qy    1618 TTGA---TAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAG 1674
            |||    | |     | |   | |||||||||||||||||||||  | |||||  ||||
Db     901 CTGACCCCAGAAAGAAGGCTGCGGTTTTAATTCTTGGAGCTGGTCGGGTCTGTCAACCAG 960

Qy    1675 CTGCTGAGTTTCTGGCATCTTACCCAGACATATGTAC---------------------CT 1713
           |||||||  |  |  |||| |    |     ||   |                       |
Db     961 CTGCTGAAATGTTATCATCATTTGGAAGGCCATCATCGAGCCAATGGTATAAAACATTGT 1020

Qy    1714 ATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATCTTTGTATCAAA 1773
             |  | ||||    |   |  |  | |  |  | || || |||| |||| |||| |  |
Db    1021 TGGAAGATGATTTTGAATGTCAAACTGATGTAGAAGTCATTGTGGGATCTCTGTACCTGA 1080

Qy    1774 AAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCAGCTTGATGTTG 1833
           | ||||||||  |||| |||| || ||| ||||  ||| |  |  |||||||||||||  
Db    1081 AGGATGCAGAGCAGACTGTTGAGGGCATTCCAAATGTAACCGGAATTCAGCTTGATGTGA 1140

Qy    1834 CTGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGC 1893
             |||  ||  |   | |   |     |||| |||||||| ||||| || || |||||||
Db    1141 TGGATCGTGCCAATTTGTGTAAGTACATTTCACAGGTTGACGTTGTTATAAGTTTGCTGC 1200

Qy    1894 CTGCTAGTTTTCATGCTGCCATTGCAGGAGTATGCATAGAGTTGAAGAAGCACATGGTAA 1953
           |  | |||| ||||  |    | |||   |  ||||| || |||| || ||  | || |
Db    1201 CCCCAAGTTGTCATATTATTGTAGCAAATGCTTGCATTGAGCTGAAAAAACATCTTGTCA 1260

Qy    1954 CGGCAAGCTATGTTGATGAATCCATGTCAAACTTGAGCCAAGCTGCCAAAGATGCAGGTG 2013
           | || |||||||||||||   ||||||||||   | |   |    || ||||||| ||  
Db    1261 CTGCTAGCTATGTTGATAGCTCCATGTCAATGCTAAATGATAAGGCTAAAGATGCTGGCA 1320

Qy    2014 TAACTATACTTTGTGAAATGGGCCTAGATCCTGGCATAGATCACTTGATGTCAATGAAGA 2073
           |||| || ||| | || ||||| | || || || || | |||  ||||| |||||||||
Db    1321 TAACAATTCTTGGAGAGATGGGCTTGGACCCAGGAATTGGTCATATGATGGCAATGAAGA 1380

Qy    2074 TGATTGATGAAGCTCATGCACGAAAGGGAAAAATAAAGGCATTTACATCTTACTGTGGTG 2133
           ||||  |  |||| ||||   | ||||| ||||||||| | || || |||| |||||||
Db    1381 TGATCAACCAAGCACATGTGAGGAAGGGGAAAATAAAGTCTTTCACTTCTTATTGTGGTG 1440

Qy    2134 GATTGCCATCTCCAGCTGCAGCAAACAATCCGCTTGCCTATAAATTCAGTTGGAACCCAG 2193
           || | |||||||| |  || || |||||||  |  | || ||||||||||||||||| || |
Db    1441 GACTTCCATCTCCTGAAGCTGCTAACAATCCATTAGCATATAAATTCAGTTGGAATCCTG 1500

Qy    2194 CTGGTGCACTCCGGTCAGGGAAAAATCCTGCAGTCTACAAATTTCTTGGTGAGACGATCC 2253
           | || ||  ||||  | |||   |||||||||   ||||||||    ||||||| ||  | |
Db    1501 CAGGAGCCATCCGAGCTGGGCGCAATCCTGCCACCTACAAATGGGGTGGTGAAACTGTAC 1560

Qy    2254 ATGTAGATGGTCATAACTTGTATGAATCAGCAAAGAGGCTCAGACTACGAGAGCTTCCAG 2313
           || | |||||  |  |  | ||||| || || |  || || || |||  ||| || ||||| |
Db    1561 ATATTGATGGGGACGATCTTTATGATTCGGCTACAAGACTAAGGCTACCGGACCTTCCTG 1620

Qy    2314 CTTTTGCTCTGGAACACTTGCCAAATCGGAATTCCTTGATATATGGTGACCTTTATGGTA 2373
           |||||||| |||||    | |||||||| ||||| ||  | ||||| ||  | ||||| |
Db    1621 CTTTTGCTTTGGAATGTCTCCCAAATCGCAATTCATTACTTTATGGGGATTTGTATGGAA 1680

Qy    2374 TCTCCAAAGAAGCATCCACCATATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTGAGA 2433
           |    |  |||||||| ||||| |   | |  || :| || || ||||| ||||||||||
Db    1681 T---AACTGAAGCATCAACCATTTTCCGTGGAACCCTCCGCTATGAAGGATTTAGTGAGA 1737

Qy    2434 TTATGGTAACCCTTTCCAAAACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGCAAG 2493
           | ||||  || || || |  | | | || ||| |   || |   ||| |  ||||     
Db    1738 TCATGGGGACACTGTCTAGGATTAGCTTATTTAACAATGAAGCCCATTCGTTGCTAATGA 1797

Qy    2494 ATACTAGTCGTCCAACATATAAGGGTTTCCTTGATGAACTACTGAATAATATCTCCACAA 2553
           ||       | ||||| |  ||    ||| |   |||||| || ||   | |       |
Db    1798 ATGGACAAAGACCAACTTTCAAAAAATTCTTATTTGAACTTCTCAAAGTTGTTGGTGATA 1857

Qy    2554 TTAACACGGACTTAGATATTGAAGCTTCTGGTGGATACGATGATGACCTGATTGCCAGAC 2613
            |      ||  ||   || | |               ||  ||||| | || |    | 
Db    1858 ATCCAGATGAACTATTGATAGGA---------------GAGAATGACATCATGGAGCAAA 1902

Qy    2614 TGTTGAAGCTCGGGTGTTGCAAAAATAAGGAAATAGCTGTTAAGACAGTCAAAACCATCA 2673
           | || |  |  |||   |||||| || |   ||  ||  |  ||||||  ||||| ||||
Db    1903 TATTAATACAAGGGCACTGCAAAGATCAAAGAACGGCAATGGAGACAGCAAAAACAATCA 1962

Qy    2674 AGTTCTTGGGACTACATGAAGAGACTCAAATACCTAAGGGTTGTTCGAGCCCATTTGATG 2733
             ||||||||||| | |||  | ||| ||||| |||      ||    ||  | |||||||
Db    1963 TTTTCTTGGGACTTCTTGACCAAACTGAAATCCCTGCTTCCTGCAAAAGTGCTTTTGATG 2022

Qy    2734 TGATTTGCCAGCGAATGGAACAGAGGATGGCCTATGGCCACAATGAGCAAGACATGGTAC 2793
           |   |||    || |||||  ||||| |  | ||   |  ||  ||  |  || ||||| |
Db    2023 TTGCTTGTTTCCGCATGGAGGAGAGGTTATCATACACCAGCACAGAAAAGGATATGGTGC 2082

Qy    2794 TGCTCCACCACGAAGTCGAGGTGGAATACCCGGACGGGCAACCCGCCGAAAAGCACCAAG 2853
           |  | || || ||||| ||  | ||||||||| ||  | ||     | || |||||   ||
Db    2083 TTTTGCATCATGAAGTGGAAATAGAATACCCAGATAGCCAAATTACAGAGAAGCATAGAG 2142

Qy    2854 CGACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCACCACTGCCATGGCGCTGA 2913
           | ||  |||| || || ||||||  |    |||  |   | ||| |||||||||||| || |
Db    2143 CTACTTTACTTGAATTTGGGAAGACTCTTGATGAAAAAACCACAACTGCCATGGCCCTTA 2202

Qy    2914 CCGTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAGGTCCAGACGA 2973
           | || || |||||||| ||  | || ||  |||| ||  |  || ||  | ||||| |
Db    2203 CTGTTGGTATTCCAGCTGCTGTTGGAGCTTTGCTTTTATTGACAAACAAAATTCAGACAA 2262

Qy    2974 AAGGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAGCATTGGAGATCTTGGAGT 3033
            ||||||  | |||||| | |||| ||| | |||  |||||||| |||| ||  | ||  
Db    2263 GAGGAGTCTTAAGGCCTATCGAACCTGAAGTATACAATCCAGCACTGGATATTATAGAAG 2322

Qy    3034 CGTCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGAAAGTTCCCTGATACACAGATAA 3093
           | |  || |||||| || | |||||    |||   |     |:     ||  |   ||| 
Db    2323 CTTATGGGATCAAGTTGATAGAGAAGACCGAGTAATTTGCATYTATGAATTGATGTATAG 2382

Qy    3094 AGATAGTATGATATAGCAGGGCACATGTATCTTTTGTATTA 3134
              ||   | || ||         |||| |  |||| || |
Db    2383 GTGTACATTAATGTACACCATGCAATGTTTGATTTGAATAA 2423
```

```
RESULT    8
AF295389
LOCUS       AF295389                1625 bp    mRNA    linear   PLN 02-SEP-2001
DEFINITION  Arabidopsis thaliana monofunctional lysine-ketoglutarate reductase
            (LKR) mRNA, complete cds.
ACCESSION   AF295389
VERSION     AF295389.1  GI:15419641
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1625)
  AUTHORS   Tang,G., Zhu,X. and Galili,G.
  TITLE     Arabidopsis monofunctional lysine-ketoglutarate reductase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1625)
  AUTHORS   Tang,G., Zhu,X. and Galili,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-2000) Plant Science, The Weizmann Institute of
            Science, Rehovot 76100, Israel
FEATURES             Location/Qualifiers
     source          1..1625
                     /organism="Arabidopsis thaliana"
```

```
                     /db_xref="taxon:3702"
     gene            1..1625
                     /gene="LKR"
     source          1..1483
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="IV"
     CDS             106..1503
                     /gene="LKR"
                     /note="results from trans-splicing event between
                     chromosomes I and IV"
                     /codon_start=1
                     /product="monofunctional lysine-ketoglutarate reductase"
                     /protein_id="AAK97099.1"
                     /db_xref="GI:15419642"
                     /translation="MNSNGHEEEKKLGNGVVGILAETVNKWERRTPLTPSHCARLLHG
                     GKDRTGISRIVVQPSAKRIHHDALYEDVGCEISDDLSDCGLILGIKQPELEMILPERA
                     YAFFSHTHKAQKENMPLLDKILSERVTLCDYELIVGDHGKRLLAFGKYAGRAGLVDFL
                     HGLGQRYLSLGYSTPFLSLGASYMYSSLAAAKAAVISVGEEIASQGLPLGICPLVFVF
                     TGTGNVSLGAQEIFKLLPHTFVEPSKLPELFVKDKGISQNGISTKRVYQVYGCIITSQ
                     DMVEHKDPSKSFDKADYYAHPEHYNPVFHEKISPYTSVLVNCMYWEKSFPCLLSTKQL
                     QDLTKKGLPLVGICDITCDIGGSIEFVNRATLIDSPFFRFNPSNNSYYDDMDGDGVLC
                     MAVDILPTEFAKEASQHFGDILSGFVGSLASMTEISDLPAHLKRACISYRGELTSLYE
                     YIPRMRKSNPEETTT"
     source          1484..1625
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="I"
BASE COUNT      514 a    301 c    339 g    471 t
ORIGIN


  Query Match             19.4%;  Score 634.6;  DB 8;  Length 1625;
  Best Local Similarity   67.2%;  Pred. No. 1.5e-152;
  Matches  914;  Conservative    0;  Mismatches  444;  Indels    3;  Gaps    1;

Qy     15 CTGCTAGGAGGAGGCAAGAACGGACCTCGAGTAAACCGGATTATTGTGCAGCCAAGCACA 74
          ||  ||   || || ||  || || |  |  |   ||| ||| | || ||||||    |
Db    226 CTTTTACACGGTGGGAAAGACAGAACCGGCATTTCCCGCATTGTGGTTCAGCCATCTGCT 285

Qy     75 AGGAGGATCCATCATGACGCTCAGTATGAGGATGCAGGATGCGAGATTTCAGAAGACCTG 134
          | | | ||||||||||| ||   |||||| ||||  || || || ||||| || ||  ||
Db    286 AAGCGTATCCATCATGATGCCTTGTATGAAGATGTTGGGTGTGAAATTTCTGATGATTTG 345

Qy    135 TCAGAATGCGGCCTTATCATAGGCATCAAACAACCCAAGCTGCAGATGATTCTTTCAGAT 194
          || || || || |||||  | || |||||||||||  ||||  | ||||||||| ||||
Db    346 TCTGATTGTGGGCTTATACTTGGAATCAAACAACCTGAGCTAGAAATGATTCTTCCAGAG 405

Qy    195 AGAGCGTACGCTTTCTTTTCACACACACACAAAGCCCAAAAAGAGAATATGCCACTGTTA 254
          ||||| ||||||||||||||||| || || || || || |||||||| |||||  ||||
Db    406 AGAGCATACGCTTTCTTTTCACATACTCATAAGGCACAGAAAGAGAACATGCCTTTGTTG 465

Qy    255 GACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAGCTAATTGTTGGAGATGAT 314
          || || || |||   || || ||| |||| |||||||||||| |||||||| ||| ||
Db    466 GATAAAATTCTTTCTGAGAGAGTGACTTTGTGTGATTATGAGCTCATTGTTGGGGATCAT 525

Qy    315 GGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCTGGACTGATAGATTTCTTA 374
          |||||| ||| | | || ||||| |||| ||| || |||||||| ||  | || ||||||
Db    526 GGGAAACGATTATTGGCGTTTGGTAAATATGCAGGCAGAGCTGGTCTTGTTGACTTCTTA 585

Qy    375 CATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACTCCATTTCTCTCTCTGGGA 434
          || || || |||||||||||| | || || |||||||| || || || ||||| || ||
Db    586 CACGGACTTGGACAGCGATATCTAAGTCTAGGATACTCAACACCTTTCCTCTCGCTCGGT 645

Qy    435 CAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCAGTCATTGTCGTTGCAGAA 494
            |||  |||||||| | ||  | ||||| || || || || || |||   ||||  |||
Db    646 GCATCGTATATGTATTCCTCATTGGCTGCTGCAAAAGCCGCTGTAATTTCTGTTGGTGAA 705

Qy    495 GAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATAGTGTTTGTGTTCACTGGA 554
          || || ||||     ||||| ||||  ||||| || ||  | || ||||| ||||| |||
Db    706 GAAATTGCAAGCCAGGGACTGCCATTAGGAATCTGCCCTCTTGTATTTGTCTTCACCGGA 765

Qy    555 GTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTATTGCCCCATACCTTTGTT 614
             ||||| || |||| ||| ||||| || || |||||| |  | || || || ||||||
Db    766 ACAGGAAATGTTTCTCTGGGGGCGCAAGAAATTTTCAAGCTTCTTCCTCACACTTTTGTT 825

Qy    615 GATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTGTCTAAGCAATCTCAG--- 671
          ||  |    || ||||| ||| | |||         ||      ||  |||  |  |
Db    826 GAACCAAGCAAACTTCCTGAACTATTTGTAAAAGACAAAGGAATTAGTCAAAATGGGATT 885

Qy    672 TCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCTAGAGACATAGTTTCT 731
          || || ||| |||| | |||| | ||||||||| || | |||     |||||| |||
Db    886 TCAACAAAGCGAGTCTATCAAGTATATGGTTGTATTATTACCAGCCAAGACATGGTTGAA 945

Qy    732 CACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCTCATCCAGAACACTAC 791
          ||||| |||||  | |    ||| |||||||  ||||||||||| || || ||||| |||
Db    946 CACAAAGATCCATCAAAGTCATTCGACAAAGCCGACTATTATGCACACCCGGAACATTAC 1005

Qy    792 ACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATCGTAAACTGTATGTAT 851
          |  || ||||| || ||||  ||  | |||||| | |||||  | ||||||||||||||
Db   1006 AATCCAGTTTTCCACGAAAAGATATCGCCATATACGTCTGTTCTTGTAAACTGTATGTAC 1065

Qy    852 TGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAGTTACAGCAATTGATGGAGACT 911
          ||||||||||| |||||     | || |  |   | ||| | ||  | || |   | |
Db   1066 TGGGAGAAGAGTTTTCCCTGTCTTCTGAGCACAAAACAGCTTCAAGATTTAACAAAAAAA 1125

Qy    912 GGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGATATTGGAGGTTCCATTGAATTT 971
          ||    ||  |||| ||| | ||||| ||||||||||| || || || ||||||||||||
Db   1126 GGACTCCCACTAGTAGGCATATGTGATATAACTTGTGACATCGGTGGCTCCATTGAATTT 1185

Qy    972 ATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGGTATGATCCTTCTAAGAATTCA 1031
           | |||     ||| | ||| ||    ||||||||  ||| | |||| || || ||||||
Db   1186 GTTAACCGAGCTACTTTAATCGATTCCCCTTTCTTCAGGTTTAATCCCTCGAACAATTCA 1245

Qy   1032 TACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTGGCTGTTGACATTCTCCCTACA 1091
          ||| | ||||| ||||| || |  || ||  | ||| |||||||||||||| | || |||
Db   1246 TACTACGATGACATGGATGGGGATGGCGTACTATGCATGGCTGTTGACATTTTACCCACA 1305

Qy   1092 GAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATACTATCTAGACTTGTTGCTAGT 1151
          |||||  | |||||||| ||||| ||||||||| | || || ||  || |||| | ||||
Db   1306 GAATTTGCAAAAGAGGCATCCCAGCATTTTGGAGATATTCTTTCCGGATTTGTCGGTAGT 1365

Qy   1152 TTGGCCTCAGTGAAGCAACCGGCAGAACTTCCTTCCTACTTGAGAAGAGCTTGCATTGCA 1211
           |||| ||| |||   ||    |||| || ||  |  |  |||  || ||||||||
Db   1366 CTGGCTTCAATGACTGAAATTTCAGATCTACCAGCACATCTGAAGAGGGCTTGCATAAGC 1425

Qy   1212 CATGCTGGCAGATTAACTCCTTTGTATGAATATATCCCTAGGATGAGAAATACTATGATA 1271
           ||   ||   ||| ||  |||||||||| |||||  ||  | ||||| ||  | |
Db   1426 TATAGGGGAGAATTGACATCTTTGTATGAGTATATTCCACGTATGAGGAAGTCAAATCCT 1485

Qy   1272 GATTTGGCACCCGCAAAAACAAATCCATTGCCTGACAAGAAGTATAGCACCCTGGTATCT 1331
          ||   | |  |  || ||||| || || |      ||      |||           | |
Db   1486 GAAGAGACTACAACATAAACAGATACAGTTTTAATCAGATTAAATATAGAGAAATGTTTT 1545

Qy   1332 CTCAGTGGGCACCTATTTGATAAGTTCCTTATAAATGAAGC 1372
           |  || ||   |  | | |||  |   |||  | | || |
Db   1546 TTTGGTTGGTCTCATTGTTATATATAAGTTAATAGTTAATC 1586


RESULT   9
BD010451
LOCUS       BD010451                 620 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Chimeric genes and methods for increasing the lysine content of the
             seeds  of plants Chimeric genes and methods for increasing the
            lysine content of the seeds  of plants.
ACCESSION   BD010451
VERSION     BD010451.1  GI:18638824
KEYWORDS    JP 2001502923-A/83.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Falco,S.C., Iii,R.E.M. and Epelbaum,S.U.
  TITLE     Chimeric genes and methods for increasing the lysine content of the
            seeds of plants
  JOURNAL   Patent: JP 2001502923-A 83 06-MAR-2001;
```

```
            EI DU PONT DE NEMOURS AND CO
COMMENT     OS   Unidentified
            PN   JP 2001502923-A/83
            PD   06-MAR-2001
            PF   27-MAR-1998 JP 1998543284
            PR   27-MAR-1997 US    08/824627
            PI   SAVERIO CARL FALCO,RAYMOND ERVIN MCDEVITT III, PI   SABINE
            URSULA EPELBAUM
            PC   C12N9/06,C12N9/12,C12N9/88,C12P13/08,C12N15/82 CC
            Strandedness: Double;
            CC   Topology: Linear;
            FH   Key              Location/Qualifiers
            FT   source           1. .620
            FT   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      162 a    128 c    141 g    189 t
ORIGIN

  Query Match             18.2%;  Score 593.2;  DB 6;  Length 620;
  Best Local Similarity   99.2%;  Pred. No. 6.7e-142;
  Matches  617;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;

Qy      237 GAGAATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAG 296
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GAGAATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAG 60

Qy      297 CTAATTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCT 356
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CTAATTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCT 120

Qy      357 GGACTGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACT 416
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db      121 GGACTGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCCACT 180

Qy      417 CCATTTCTCTCTCTGGACAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCA 476
            |||||||||||||||||| ||||||||||| |||||||||||||||||||||||||||||
Db      181 CCATTTCTCTCTCTGGAC-ATCTCATATGT-TCCTTCGCTCGCTGCAGCCAAGGCTGCA 238

Qy      477 GTCATTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATA 536
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      239 GTCATTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATA 298

Qy      537 GTGTTTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTA 596
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      299 GTGTTTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTA 358

Qy      597 TTGCCCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTG 656
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      359 TTGCCCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTG 418

Qy      657 TCTAAGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCT 716
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      419 TCTAAGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCT 478

Qy      717 AGAGACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCT 776
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      479 AGAGACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCT 538

Qy      777 CATCCAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATC 836
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      539 CATCCAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATC 598

Qy      837 GTAAACTGTATGTATTGGGAGA 858
            |||||||| ||||||||||| |
Db      599 GTAAACTGCATGTATTGGGAAA 620

RESULT  10
AY039906
```

```
LOCUS       AY039906               1905 bp    mRNA    linear   PLN 24-JUN-2001
DEFINITION  Arabidopsis thaliana AT4g33150/F4I10_80 mRNA, complete cds.
ACCESSION   AY039906
VERSION     AY039906.1  GI:14532563
KEYWORDS    FLI_CDNA.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1905)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J.,
            Bowser,L., Carninci,P., Dale,J.M., Gibson,H.A., Goldsmith,A.D.,
            Hayashizaki,Y., Ishida,J., Jiang,P.X., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X.,
            Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J.,
            Pham,P.K., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A.,
            Tang,C.C., Toriumi,M., Yamada,K., Yu,G., Yu,S., Shinozaki,K.,
            Davis,R.W., Theologis,A. and Ecker,J.R.
  TITLE     Arabidopsis cDNA clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1905)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J.,
            Bowser,L., Carninci,P., Dale,J.M., Gibson,H.A., Goldsmith,A.D.,
            Hayashizaki,Y., Ishida,J., Jiang,P.X., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X.,
            Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J.,
            Pham,P.K., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A.,
            Tang,C.C., Toriumi,M., Yamada,K., Yu,G., Yu,S., Shinozaki,K.,
            Davis,R.W., Theologis,A. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2001) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H.,
            Kim,C.J., Koesema,E., Meyers,M.C., Shinn,P., Banh,J. Bowser,L.,
            Dale,J.M., Gibson,H.A., Goldsmith,A.D., Jiang,P.X., Jones,T.,
            Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
            Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K., Quach,H.L.,
            Southwick,A., Tang,C.C., Toriumi,M., Yamada,K., Yu,G., Yu,S.,
            Davis,R.W.,  Theologis,A., and Ecker,J.R.

            Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .1905
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /clone="RAFL05-02-O17(R12911)"
                     /note="ecotype: Columbia"
     5'UTR           1. .164
     CDS             165. .1613
                     /note="lysine-ketoglutarate reductase/saccharopine"
                     /codon_start=1
                     /product="AT4g33150/F4I10_80"
                     /protein_id="AAK64010.1"
                     /db_xref="GI:14532564"
                     /translation="MTKKSGVLILGAGRVCRPAADFLASVRTISSQQWYKTYVGADSE
                     EKTDVHVIVASLYLKDAKETVEGISDVEAVRLDVSDSESLLKYVSQVDVVLSLLPASC
                     HAVVAKTCIELKKHLVTASYVDDETSMLHEKAKSAGITILGEMGLDPGIDHMMAMKMI
                     NDAHIKKGKVKSFTSYCGGLPSPAAANNPLAYKFSWNPAGAIRAGQNPAKYKSNGDII
                     HVDGKNLYDSAARFRVPNLPAFALECFPNRDSLVYGEHYGIESEATTIFRGTLRYEGF
                     SMIMATLSKLGFFDSEANQVLSTGKRITFGALLSNILNKDADNESEPLAGEEEISKRI
```

```
                     IKLGHSKETAAKAAKTIVFLGFNEEREVPSLCKSVFDATCYLMEEKLAYSGNEQDMVL
                     LHHEVEVEFLESKRIEKHTATLLEFGDIKNGQTTTAMAKTVGIPAAIGALLLIEDKIK
                     TRGVLRPLEAEVYLPALDILQAYGIKLMEKAE"
     3'UTR           1614..1905
BASE COUNT      634 a    328 c    449 g    494 t
ORIGIN


  Query Match             16.5%;  Score 537.8;  DB 8;  Length 1905;
  Best Local Similarity   60.5%;  Pred. No. 1.5e-127;
  Matches  997;  Conservative    1;  Mismatches  598;  Indels   51;  Gaps    5;

Qy     1465 AAGTAGGAGCAGATGATACTGCCACATTGGATAAAATTATTGATTCCTTGACTTCTTTAG 1524
            | || || || |||||||       ||||||| |||| |||||||| || |||   ||||
Db        1 AGGTTGGTGCGGATGATAAGAGAGTATTGGATCAAATCATTGATTCATTAACTCGGTTAG 60

Qy     1525 CTAATGAACATGGTGGAGATCACGATGCCG---GGCAAGAAATTGAATTAGCTCTGAAGA 1581
            |||||  | |||  |   ||       |     |  ||| || | |  |  | |||||||
Db       61 CTAATCCAAATGAAGATTATATATCCCCACATAGAGAAGCAAATAAGATCTCACTGAAGA 120

Qy     1582 TAGGAAAAGTCAATGAGTATGA------AACTGACGTCACAATTGATAAAGGAGGGCCAA 1635
            | || |||||| |  |  |  |      ||  ||     |       |    |    ||
Db      121 TTGGTAAAGTCCAGCAAGAAAATGAGATAAAAGAGAAGCCTGAAATGACGAAAAAATCAG 180

Qy     1636 AGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTGAGTTTCTGGCATCTT 1695
               |||| ||||||||| |||||| | || ||||| |||||||||||| || || || ||
Db      181 GTGTTTTGATTCTTGGTGCTGGACGTGTGTGTCGCCCAGCTGCTGATTTCCTAGCTTCAG 240

Qy     1696 ACCCAGACATATGTACCTAT---------------------GGTGTTGATGACCATGATG 1734
                |  ||| |   |  |                      || |  ||     | ||
Db      241 TTAGAACCATTTCGTCACAGCAATGGTACAAAACATATGTCGGAGCAGACTCTGAAGAGA 300

Qy     1735 CAGATCAAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTG 1794
             |    |  ||||||| || || || ||| ||||||  || |||||  ||||| ||||
Db      301 AAACAGATGTTCATGTGATTGTCGCGTCTCTGTATCTTAAGGATGCCAAAGAGACGGTTG 360

Qy     1795 ATGGTATTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAG 1854
            | ||||||  | ||  |   ||    | || | |||||  |||||| || ||| ||
Db      361 AAGGTATTTCAGATGTAGAAGCAGTTCGGCTAGATGTATCTGATAGTGAAAGTCTCCTTA 420

Qy     1855 ATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCA 1914
            |   ||||||||||||||| |||||  | || ||  | |||||| |||| |||||||
Db      421 AGTATGTTTCTCAGGTTGATGTTGTCCTAAGTTTATTACCTGCAAGTTGTCATGCTGTTG 480

Qy     1915 TTGCAGGAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAAT 1974
            | |||     |||||| ||| ||||||||||  | || || || ||||||||||||||
Db      481 TAGCAAAGACATGCATTGAGCTGAAGAAGCATCTCGTCACTGCTAGCTATGTTGATGATG 540

Qy     1975 CCATGTCAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGG 2034
              | ||| |  ||     |    || ||   ||| ||  ||||| || ||  | |||||||
Db      541 AAACGTCCATGTTACATGAGAAGGCTAAGAGTGCTGGGATAACGATTCTAGGCGAAATGG 600

Qy     2035 GCCTAGATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCAC 2094
            | || || ||||||    |||||| ||||| | ||||| |||||  | || ||||||
Db      601 GACTGGACCCTGGAATCGATCACATGATGGCGATGAAAATGATCAACGATGCTCATATCA 660

Qy     2095 GAAAGGGAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAG 2154
             ||| || ||| | ||| | ||||| ||||| ||||| ||  | || ||||| |||||||
Db      661 AAAAAGGGAAAGTGAAGTCTTTTACCTCTTATTGTGGAGGGCTTCCCTCTCCTGCTGCAG 720

Qy     2155 CAAACAATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGA 2214
            |||| |||||  | || |||||||| || ||||||||| ||| ||| | ||  | ||
Db      721 CAAATAATCCATTAGCATATAAATTTAGCTGGAACCCTGCTGGAGCAATTCGAGCTGGTC 780

Qy     2215 AAAATCCTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGT 2274
            |||| || ||    ||||||      || || |  || ||||| |||||  |  | |  |
Db      781 AAAACCCCGCCAAATACAAAAGCAACGGCGACATAATACATGTTGATGGGAAGAATCTCT 840

Qy     2275 ATGAATCAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGC 2334
            |||| || ||    ||  || || |||   | ||||||||||||||||  ||||    || |
Db      841 ATGATTCCGCGGCAAGATTCCGAGTACCTAATCTTCCAGCTTTTGCATTGGAGTGTTTTC 900

Qy     2335 CAAATCGGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCA 2394
            |||||||  | |||||| | || || || | |||||| |||   |  ||||| | || |
Db      901 CAAATCGTGACTCCTTGGTTTACGGGGAACATTATGGCATCGAGAGCGAAGCAACAACGA 960

Qy     2395 TATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAA 2454
            ||| | | |  || :|  | || ||||| |||||  ||| |||| ||| ||||| |||
Db      961 TATTTCGTGGAACACTCAGATATGAAGGGTTTAGTATGATAATGGCAACACTTTCGAAAC 1020

Qy     2455 CTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATA 2514
             ||| ||||||||   || |    ||| | | ||      |   |   |    || | |
Db     1021 TTGGATTCTTTGACAGTGAAGCAAATCAAGTACTCTCCACTGGAAAGAGGATTACGTTTG 1080

Qy     2515 AGGGTTTCCTTGATGAACTACTGAATAATATCTCCACAATTAACACGGACTTAGATATTG 2574
              | | |  |   | |  | || |||||     |    | | |  |
Db     1081 GTGCTCTTTTAAGTAACATTCTAAATAAGGATGCAGACAATGAATC-------------- 1126

Qy     2575 AAGCTTCTGGTGGATACGATGATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTTGCA 2634
             ||   |    |     || || ||  | |     ||| |  | ||||| ||   || |
Db     1127 -AGAGCCCCTAGCGGGAGAAGAAGAGATAAGCAAGAGAATTATCAAGCTTGGACATTCC- 1184

Qy     2635 AAAATAAGGAAATAGCTGTTAAGACAGTCAAAACCATCAAGTTCTTGGGACTACATGAAG 2694
                 ||||| |  || |  ||  | | |||||| ||    ||||||||  |  | ||||
Db     1185 -----AAGGAGACTGCAGCCAAAGCTGCCAAAACAATTGTATTCTTGGGGTTCAACGAAG 1239

Qy     2695 AGACTCAAATACCTAAGGGTTGTTCGAGCCCATTTGATGTGATTTGCCAGCGAATGGAAC 2754
            |||   |  | ||       |||   |||  |||||||| | |||  | | |||||||
Db     1240 AGAGGGAGGTTCCATCACTGTGTAAAAGCGTATTTGATGCAACTTGTTACCTAATGGAAG 1299

Qy     2755 AGAGGATGGCCTATGGCCACAATGAGCAAGACATGGTACTGCTCCACCACGAAGTCGAGG 2814
            |||   | || |||  |   ||||| || |||||||| ||  | || |||||||| || |
Db     1300 AGAAACTAGCTTATTCCGGAAATGAACAGGACATGGTGCTTTTGCATCACGAAGTAGAAG 1359

Qy     2815 TGGAATACCCGGACGGGCAACCCGCCGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGA 2874
            |||||| ||  ||  |  |||     || ||||||   ||||| ||  |||| ||||||
Db     1360 TGGAATTCCTTGAAAGCAAACGTATAGAGAAGCACACTGCGACTCTTTTGGAATTCGGGG 1419

Qy     2875 AGGTTGAAAATGGCAGGTCCACCACTGCCATGGCGCTGACCGTCGGCATTCCAGCAGCAA 2934
            |  |  | |||||     | || || || |||||   ||| || || || || ||||| |
Db     1420 ACATCAAGAATGGACAAACAACAACCGCTATGGCCAAGACTGTTGGGATCCCTGCAGCCA 1479

Qy     2935 TAGGGGCCCTGCTATTGCTAAAGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGC 2994
            | || || ||||| ||  |  |  | |||  || |||| | ||||||  | ||||||||
Db     1480 TTGGAGCTCTGCTGTTAATTGAAGACAAGATCAAGACAAGAGGAGTCTTAAGGCCTCTCG 1539

Qy     2995 AACCGGAAATCTACGTTCCAGCATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTG 3054
            || | ||  | ||  | |||||| ||||| || ||| |  | |  || || |||||| | |
Db     1540 AAGCAGAGGTGTATTTGCCAGCTTTGGATATATTGCAAGCATATGGTATAAAGCTGATGG 1599

Qy     3055 AGAAAGTGGAGACTTGAAAGTTCCCTG 3081
            |||| |  ||    | ||||  | |||
Db     1600 AGAAGGCAGAATGATCAAAGAACTCTG 1626


RESULT 11
BD010455
LOCUS       BD010455                 720 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Chimeric genes and methods for increasing the lysine content of the
             seeds  of plants Chimeric genes and methods for increasing the
            lysine content of the seeds  of plants.
ACCESSION   BD010455
VERSION     BD010455.1  GI:18638828
KEYWORDS    JP 2001502923-A/87.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Falco,S.C., Iii,R.E.M. and Epelbaum,S.U.
  TITLE     Chimeric genes and methods for increasing the lysine content of the
            seeds of plants
  JOURNAL   Patent: JP 2001502923-A 87 06-MAR-2001;
            EI DU PONT DE NEMOURS AND CO
COMMENT     OS   Oryza sativa (rice)
```

```
            PN    JP 2001502923-A/87
            PD    06-MAR-2001
            PF    27-MAR-1998 JP 1998543284
            PR    27-MAR-1997 US     08/824627
            PI    SAVERIO CARL FALCO,RAYMOND ERVIN MCDEVITT III, PI   SABINE
            URSULA EPELBAUM
            PC    C12N9/06,C12N9/12,C12N9/88,C12P13/08,C12N15/82 CC
            Strandedness: Single;
            CC    Topology: Linear;
            FH    Key               Location/Qualifiers
            FT    CDS               2. .720.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
BASE COUNT      203 a    150 c    158 g    206 t      3 others
ORIGIN

  Query Match             16.0%;  Score 522.6;  DB 6;  Length 720;
  Best Local Similarity   85.5%;  Pred. No. 1.1e-123;
  Matches  579;  Conservative    1;  Mismatches   97;  Indels    0;  Gaps    0;

Qy   1848 CTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCAT 1907
          |||||  ||||||||||||||||||| || ||| |||||||||||||||| |||||||||
Db     11 CTTTCCAATCTTGTTTCTCAGGTTGAAGTAGTAGTTAGCTTGCTGCCTGCCAGTTTTCAT 70

Qy   1908 GCTGCCATTGCAGGAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTT 1967
          |||||||| ||| |||||||||||||| ||||||||||| |||| || ||||||||||||
Db     71 GCTGCCATAGCAAGAGTATGCATAGAGATGAAGAAGCACTTGGTCACTGCAAGCTATGTT 130

Qy   1968 GATGAATCCATGTCAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGT 2027
          ||||| ||||||||||| |||   ||| ||||  ||| ||| ||||||||||| || |||
Db    131 GATGAGTCCATGTCAAAGTTGGAACAATCTGCAGAAGGTGCTGGTGTAACTATTCTCTGT 190

Qy   2028 GAAATGGGCCTAGATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCT 2087
          ||||||||||| |||||||||||| ||||  |||||||||||||||||||||| |||||
Db    191 GAAATGGGCCTGGATCCTGGCATANATCATATGATGTCAATGAAGATGATTGACGAAGCA 250

Qy   2088 CATGCACGAAAGGGAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCA 2147
          ||| |||| ||||| ||||||||| |||||||||| |  ||||| ||| | |||||||||
Db    251 CATTCACGGAAGGGGAAAATAAAGTCATTTACATCCTTTTGTGGAGGACTTCCATCTCCA 310

Qy   2148 GCTGCAGCAAACAATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGG 2207
          ||| | ||||||||||| ||||| ||||| ||||||||||  |||||||||||  ||||
Db    311 GCTTCTGCAAACAATCCACTTGCTTATAAGTTCAGTTGGAGTCCAGCTGGTGCCATCCGT 370

Qy   2208 TCAGGGAAAAATCCTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCAT 2267
           |||||| ||| ||||| ||||||||||||| ||| || |  ||||||||||||||| ||
Db    371 GCAGGGAGAAACCCTGCTGTCTACAAATTTCATGGAGAAATCATCCATGTAGATGGTGAT 430

Qy   2268 AACTTGTATGAATCAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAA 2327
          || ||||||||||| ||||||||||||||| ||| ||| |||||||||||||| ||||||
Db    431 AAATTGTATGAATCCGCAAAGAGGCTCAGATTACMAGAACTTCCAGCTTTTGCACTGGAA 490

Qy   2328 CACTTGCCAAATCGGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCA 2387
          ||||||||||| |||||||||||||| ||||| ||||| ||||| |||||||||||||||
Db    491 CACTTGCCAAACCGGAATTCCTTGATGTATGGAGACCTGTATGGGATCTCCAAAGAAGCA 550

Qy   2388 TCCACCATATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTT 2447
          || ||  | || |||||||||:||||||| ||||| |||| |||||| |||| |||| |
Db    551 TCTACTGTGTACAGGGCTACTCTTCGTTATGAAGGATTTAATGAGATAATGGCAACCTTC 610

Qy   2448 TCCAAAACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCA 2507
           | |||| |||||| ||||||||||||| |||||||||| ||||| | |||| ||| ||
Db    611 GCGAAAATTGGGTTTTTTGATGCTGCAAGTCATCCACTGTTGCAACAAACTACTCGCCCT 670

Qy   2508 ACATATAAGGGTTTCCT 2524
          ||||| | || ||||||
Db    671 ACATACANGGATTTCCT 687

RESULT 12
```

```
BD010458
LOCUS       BD010458                1449 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Chimeric genes and methods for increasing the lysine content of the
             seeds  of plants Chimeric genes and methods for increasing the
            lysine content of the seeds  of plants.
ACCESSION   BD010458
VERSION     BD010458.1  GI:18638831
KEYWORDS    JP 2001502923-A/90.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1449)
  AUTHORS   Falco,S.C., Iii,R.E.M. and Epelbaum,S.U.
  TITLE     Chimeric genes and methods for increasing the lysine content of the
            seeds of plants
  JOURNAL   Patent: JP 2001502923-A 90 06-MAR-2001;
            EI DU PONT DE NEMOURS AND CO
COMMENT     OS   Unidentified
            PN   JP 2001502923-A/90
            PD   06-MAR-2001
            PF   27-MAR-1998 JP 1998543284
            PR   27-MAR-1997 US     08/824627
            PI   SAVERIO CARL FALCO,RAYMOND ERVIN MCDEVITT III, PI   SABINE
            URSULA EPELBAUM
            PC   C12N9/06,C12N9/12,C12N9/88,C12P13/08,C12N15/82 CC
            Strandedness: Double;
            CC   Topology: Linear;
            FH   Key               Location/Qualifiers
            FT   source            1. .1449
            FT   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1449
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      446 a    267 c    356 g    380 t
ORIGIN

  Query Match             15.8%;  Score 515;  DB 6;  Length 1449;
  Best Local Similarity   61.7%;  Pred. No. 1.1e-121;
  Matches  893;  Conservative    1;  Mismatches  511;  Indels   42;  Gaps    3;

Qy   1639 TTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTGAGTTTCTGGCATCTTACC 1698
          |||| |||||||| |||||| | || ||||| ||||||||||| || || || ||
Db     20 TTTTGATTCTTGGTGCTGGACGTGTGTGTCGCCCAGCTGCTGATTTCCTAGCTTCAGTTA 79

Qy   1699 CAGACATATGTACCTAT---------------------GGTGTTGATGACCATGATGCAG 1737
           |  ||| |   |  |                      || |  ||     | ||   |
Db     80 GAACCATTTCGTCACAGCAATGGTACAAAACATATTTCGGAGCAGACTCTGAAGAGAAAA 139

Qy   1738 ATCAAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATG 1797
             |  ||||||| || || || ||| ||||||  || |||||  ||||||| ||||| |
Db    140 CAGATGTTCATGTGATTGTCGCGTCTCTGTATCTTAAGGATGCCAAAGAGACGGTTGAAG 199

Qy   1798 GTATTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATC 1857
          |||||  | ||  |   ||    | ||| |||||  |||||| || ||| ||     |
Db    200 GTATTTCAGATGTAGAAGCAGTTCGGCTAGATGTATCTGATAGTGAAAGTCTCCTTAAGT 259

Qy   1858 TTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTG 1917
           ||||||||||||||| |||||  | || ||  | ||||| |||| ||||||||   | |
Db    260 ATGTTTCTCAGGTTGATGTTGTCCTAAGTTTATTACCTGCAAGTTGTCATGCTGTTGTAG 319

Qy   1918 CAGGAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCA 1977
          ||     |||||| ||| ||||||||||  | || || || ||||||||||||||    |
Db    320 CAAAGACATGCATTGAGCTGAAGAAGCATCTCGTCACTGCTAGCTATGTTGATGATGAAA 379

Qy   1978 TGTCAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCC 2037
           ||| |  ||     |    || ||   ||| ||  |||| || ||  | |||||||| |
Db    380 CGTCCATGTTACATGAGAAGGCTAAGAGTGCTGGGATAACGATTCTAGGCGAAATGGGAC 439

Qy   2038 TAGATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAA 2097
          | || ||||| || |||||| ||||| | ||||| |||||  | || ||||||     ||
Db    440 TGGACCCTGGAATCGATCACATGATGGCGATGAAAATGATCAACGATGCTCATATCAAAA 499
```

```
Qy   2098 AGGGAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAA 2157
          | || ||| | ||| | ||||| ||||| ||||| ||  | || ||||| ||||||||||
Db    500 AAGGGAAAGTGAAGTCTTTTACCTCTTATTGTGGAGGGCTTCCCTCTCCTGCTGCAGCAA 559

Qy   2158 ACAATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAA 2217
          | |||||  | || |||||||| || |||||||| ||||| ||| | ||  | ||  |||
Db    560 ATAATCCATTAGCATATAAATTTAGCTGGAACCCTGCTGGAGCAATTCGAGCTGGTCAAA 619

Qy   2218 ATCCTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATG 2277
          | || ||    ||||||      || || |  || ||||| |||||  | ||  | ||||
Db    620 ACCCCGCCAAATACAAAAGCAACGGCGACATAATACATGTTGATGGGAAGAATCTCTATG 679

Qy   2278 AATCAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAA 2337
          | || ||    ||  || || |||   | |||||||||||||||  ||||    || ||||
Db    680 ATTCCGCGGCAAGATTCCGAGTACCTAATCTTCCAGCTTTTGCATTGGAGTGTTTTCCAA 739

Qy   2338 ATCGGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATAT 2397
          ||||  | ||||||| | || || || | ||||||| |||   |  |||||| | || ||||
Db    740 ATCGTGACTCCTTGGTTTACGGGGAACATTATGGCATCGAGAGCGAAGCAACAACGATAT 799

Qy   2398 ATAGGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTG 2457
           | | |  || :|  | || ||||| ||||||  ||| |||| ||| ||||| |||  ||
Db    800 TTCGTGGAACACTCAGATATGAAGGGTTTAGTATGATAATGGCAACACTTTCGAAACTTG 859

Qy   2458 GGTTCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATAAGG 2517
          | ||||||||   || |     ||| | | ||      |    |   |    || | |   |
Db    860 GATTCTTTGACAGTGAAGCAAATCAAGTACTCTCCACTGGAAAGAGGATTACGTTTGGTG 919

Qy   2518 GTTTCCTTGATGAACTACTGAATAATATCTCCACAATTAACACGGACTTAGATATTGAAG 2577
           | |  |   | |  | || |||||     |    | | |  |                ||
Db    920 CTCTTTTAAGTAACATTCTAAATAAGGATGCAGACAATGAATC---------------AG 964

Qy   2578 CTTCTGGTGGATACGATGATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTTGCAAAA 2637
             |    |     || || ||  | |     ||| |  | ||||| ||   || |
Db    965 AGCCCCTAGCGGGAGAAGAAGAGATAAGCAAGAGAATTATCAAGCTTGGACATTCC---- 1020

Qy   2638 ATAAGGAAATAGCTGTTAAGACAGTCAAAACCATCAAGTTCTTGGGACTACATGAAGAGA 2697
            ||||| |  || |  ||  | | ||||||| ||    ||||||||  |  | |||||||
Db   1021 --AAGGAGACTGCAGCCAAAGCTGCCAAAACAATTGTATTCTTGGGGTTCAACGAAGAGA 1078

Qy   2698 CTCAAATACCTAAGGGTTGTTCGAGCCCATTTGATGTGATTTGCCAGCGAATGGAACAGA 2757
             |  | ||       |||   |||  ||||||||  | |||  | | |||||||| |||
Db   1079 GGGAGGTTCCATCACTGTGTAAAAGCGTATTTGATGCAACTTGTTACCTAATGGAAGAGA 1138

Qy   2758 GGATGGCCTATGGCCACAATGAGCAAGACATGGTACTGCTCCACCACGAAGTCGAGGTGG 2817
             | || ||| |  |   |||||| || |||||||| ||  | || |||||||| || ||||
Db   1139 AACTAGCTTATTCCGGAAATGAACAGGACATGGTGCTTTTGCATCACGAAGTAGAAGTGG 1198

Qy   2818 AATACCCGGACGGGCAACCCGCCGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGAAGG 2877
          ||| ||  ||  |  |||     || ||||||   ||||| ||  |||| |||||| |
Db   1199 AATTCCTTGAAAGCAAACGTATAGAGAAGCACACTGCGACTCTTTTGGAATTCGGGGACA 1258

Qy   2878 TTGAAAATGGCAGGTCCACCACTGCCATGGCGCTGACCGTCGGCATTCCAGCAGCAATAG 2937
          |  | |||||     | || || || |||||   ||| || || || || ||||| || |
Db   1259 TCAAGAATGGACAAACAACAACCGCTATGGCCAAGACTGTTGGGATCCCTGCAGCCATTG 1318

Qy   2938 GGGCCCTGCTATTGCTAAAGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGCAAC 2997
          | || ||| | ||  |  |  | ||| || |||| | ||||||  | ||||||||  ||
Db   1319 GAGCTCTGGTGTTAATTGAAGACAAGATCAAGACAAGAGGAGTCTTAAGGCCTCTCGAAG 1378

Qy   2998 CGGAAATCTACGTTCCAGCATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGA 3057
          | ||  | ||  | ||||| ||||| || ||| |  | |  || || ||| | ||||
Db   1379 CAGAGGTGTATTTGCCAGCTTTGGATATATTGCAAGCATATGGTATAAAGCTGATGGAGA 1438

Qy   3058 AAGTGGA 3064
          | |  ||
Db   1439 AGGCAGA 1445


RESULT  13
ATU90523
LOCUS       ATU90523                1596 bp    mRNA    linear   PLN 27-SEP-2001
```

```
DEFINITION  Arabidopsis thaliana monofunctional saccharopine dehydrogenase
            (SDH) mRNA, complete cds.
ACCESSION   U90523
VERSION     U90523.1  GI:4099846
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1596)
  AUTHORS   Tang,G., Miron,D., Zhu-Shimoni,J.X. and Galili,G.
  TITLE     Regulation of lysine catabolism through lysine-ketoglutarate
            reductase and saccharopine dehydrogenase in Arabidopsis
  JOURNAL   Plant Cell 9 (8), 1305-1316 (1997)
  MEDLINE   97432144
   PUBMED   9286108
REFERENCE   2  (bases 1 to 1596)
  AUTHORS   Tang,G., Miron,D., Zhu-Shimoni,J.X. and Galili,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1997) Dept. of Plant Genetics, The Weizmann
            Institute of Science, Rehovot 76100, Israel
FEATURES             Location/Qualifiers
     source          1. .1596
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
     gene            1. .1596
                     /gene="SDH"
     CDS             52. .1500
                     /gene="SDH"
                     /EC_number="1.5.1.9"
                     /function="catalyzes the second step in lysine catabolism"
                     /note="monofunctional enzyme"
                     /codon_start=1
                     /product="saccharopine dehydrogenase"
                     /protein_id="AAD00700.1"
                     /db_xref="GI:4099847"
                     /translation="MTKKSGVLILGAGRVCRPAADFLASVRTISSQQWYKTYFGADSE
                     EKTDVHVIVASLYLKDAKETVEGISDVEAVRLDVSDSESLLKYVSQVDVVLSLLPASC
                     HAVVAKTCIELKKHLVTASYVDDETSMLHENAKSAGITILGEMGLDPGIDPMMAKKMI
                     NDAPITKGKVKSFTSYCGGLPSRAAANNPLAYKFSWNPAGAIRAGQNPAKYKSNGDII
                     HVDGKNLYDSAARFRVPNLPAFALECFPNRDSLVYGEHYGIESEATTIFRGTLRYEGF
                     SMIMATLSKLGFFDSEANQVLSTGKRITFGALLSNILNKDADNESEPLAGEEEISKRI
                     IKLGHSKETAAKAAKTIVFLGFNEEREVPSLCKSVFDATCYLMEEKLAYSGNEQDMVL
                     LHHEVEVEFLESKRIEKHTATLLEFGDIKNGQTTTAMAKTVGIPAAIGALLLIEDKIK
                     TRGVLRPLEAEVYLPALDILQAYGIKLMEKAE"
BASE COUNT      512 a    291 c    385 g    408 t
ORIGIN


  Query Match             15.8%;  Score 514.4;  DB 8;  Length 1596;
  Best Local Similarity   61.4%;  Pred. No. 1.6e-121;
  Matches  899;  Conservative    1;  Mismatches  522;  Indels   42;  Gaps    3;

Qy   1639 TTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTGAGTTTCTGGCATCTTACC 1698
          |||| |||||||| |||||| | || ||||| ||||||||||| || || || ||
Db     71 TTTTGATTCTTGGTGCTGGACGTGTGTGTCGCCCAGCTGCTGATTTCCTAGCTTCAGTTA 130

Qy   1699 CAGACATATGTACCTAT---------------------GGTGTTGATGACCATGATGCAG 1737
           |  ||| |   |  |                      || |  ||     | ||   |
Db    131 GAACCATTTCGTCACAGCAATGGTACAAAACATATTTCGGAGCAGACTCTGAAGAGAAAA 190

Qy   1738 ATCAAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATG 1797
             |  |||||||| || || ||| ||||||  || |||||  ||||||| ||||| |
Db    191 CAGATGTTCATGTGATTGTCGCGTCTCTGTATCTTAAGGATGCCAAAGAGACGGTTGAAG 250

Qy   1798 GTATTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATC 1857
          |||||  | ||  |   ||    | ||| ||||||  |||||| || ||| ||     |
Db    251 GTATTTCAGATGTAGAAGCAGTTCGGCTAGATGTATCTGATAGTGAAAGTCTCCTTAAGT 310

Qy   1858 TTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTG 1917
           ||||||||||||||| |||||  | || ||  | ||||| |||| ||||||||   | |
Db    311 ATGTTTCTCAGGTTGATGTTGTCCTAAGTTTATTACCTGCAAGTTGTCATGCTGTTGTAG 370
```

```
Qy   1918 CAGGAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCA 1977
          ||     |||||| ||| ||||||||||  | || || || ||||||||||||||    |
Db    371 CAAAGACATGCATTGAGCTGAAGAAGCATCTCGTCACTGCTAGCTATGTTGATGATGAAA 430

Qy   1978 TGTCAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCC 2037
           ||| |  ||     |    || ||   ||| ||  |||| || ||  | |||||||| |
Db    431 CGTCCATGTTACATGAGAACGCTAAGAGTGCTGGGATAACGATTCTAGGCGAAATGGGAC 490

Qy   2038 TAGATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAA 2097
          | || ||||| || |||| | ||||| | | ||| |||||  | || |||| |     ||
Db    491 TGGACCCTGGAATCGATCCCATGATGGCGAAGAAAATGATCAACGATGCTCCTATCACAA 550

Qy   2098 AGGGAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAA 2157
          | || ||| | ||| | ||||| ||||| ||||| ||  | || ||||  ||||||||||
Db    551 AAGGGAAAGTGAAGTCTTTTACCTCTTATTGTGGAGGGCTTCCCTCTCGTGCTGCAGCAA 610

Qy   2158 ACAATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAA 2217
          | |||||  | || |||||||| || |||||||| ||||| ||| | ||  | ||  |||
Db    611 ATAATCCATTAGCATATAAATTTAGCTGGAACCCTGCTGGAGCAATTCGAGCTGGTCAAA 670

Qy   2218 ATCCTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATG 2277
          | || ||    ||||||      || || |  || ||||| |||||  | ||  | ||||
Db    671 ACCCCGCCAAATACAAAAGCAACGGCGACATAATACATGTTGATGGGAAGAATCTCTATG 730

Qy   2278 AATCAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAA 2337
          | || ||    ||  || ||  |||   | ||||||||||||||  ||||    || ||||
Db    731 ATTCCGCGGCAAGATTCCGAGTACCTAATCTTCCAGCTTTTGCATTGGAGTGTTTTCCAA 790

Qy   2338 ATCGGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATAT 2397
          ||||  | |||||| | || || || | |||||| |||   |  |||||| | || ||||
Db    791 ATCGTGACTCCTTGGTTTACGGGGAACATTATGGCATCGAGAGCGAAGCAACAACGATAT 850

Qy   2398 ATAGGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTG 2457
           | | |  || :|  | || ||||| ||||||  ||| |||| ||| ||||| |||  ||
Db    851 TTCGTGGAACACTCAGATATGAAGGGTTTAGTATGATAATGGCAACACTTTCGAAACTTG 910

Qy   2458 GGTTCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATAAGG 2517
          | ||||||||   || |    ||| | | ||      |   |   |    || | |   |
Db    911 GATTCTTTGACAGTGAAGCAAATCAAGTACTCTCCACTGGAAAGAGGATTACGTTTGGTG 970

Qy   2518 GTTTCCTTGATGAACTACTGAATAATATCTCCACAATTAACACGGACTTAGATATTGAAG 2577
           | |  |   | |  | || |||||     |    | | |  |               ||
Db    971 CTCTTTTAAGTAACATTCTAAATAAGGATGCAGACAATGAATC---------------AG 1015

Qy   2578 CTTCTGGTGGATACGATGATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTTGCAAAA 2637
             |    |     || || ||  | |     ||| |  | ||||| ||   || |
Db   1016 AGCCCCTAGCGGGAGAAGAAGAGATAAGCAAGAGAATTATCAAGCTTGGACATTCC---- 1071

Qy   2638 ATAAGGAAATAGCTGTTAAGACAGTCAAAACCATCAAGTTCTTGGGACTACATGAAGAGA 2697
            ||||| |  || |  ||  | | |||||| ||    ||||||||  |  | |||||||
Db   1072 --AAGGAGACTGCAGCCAAAGCTGCCAAAACAATTGTATTCTTGGGGTTCAACGAAGAGA 1129

Qy   2698 CTCAAATACCTAAGGGTTGTTCGAGCCCATTTGATGTGATTTGCCAGCGAATGGAACAGA 2757
             |  | ||       |||   |||  ||||||||  | |||  | | ||||||| |||
Db   1130 GGGAGGTTCCATCACTGTGTAAAAGCGTATTTGATGCAACTTGTTACCTAATGGAAGAGA 1189

Qy   2758 GGATGGCCTATGGCCACAATGAGCAAGACATGGTACTGCTCCACCACGAAGTCGAGGTGG 2817
             | || |||  |   ||||| || |||||||| ||  | || |||||||| || ||||
Db   1190 AACTAGCTTATTCCGGAAATGAACAGGACATGGTGCTTTTGCATCACGAAGTAGAAGTGG 1249

Qy   2818 AATACCCGGACGGGCAACCCGCCGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGAAGG 2877
          ||| ||  ||  |  |||     || ||||||   ||||| ||  |||| |||||| |
Db   1250 AATTCCTTGAAAGCAAACGTATAGAGAAGCACACTGCGACTCTTTTGGAATTCGGGGACA 1309

Qy   2878 TTGAAAATGGCAGGTCCACCACTGCCATGGCGCTGACCGTCGGCATTCCAGCAGCAATAG 2937
          |  | |||||     | || || || |||||   ||| || || || || ||||| || |
Db   1310 TCAAGAATGGACAAACAACAACCGCTATGGCCAAGACTGTTGGGATCCCTGCAGCCATTG 1369

Qy   2938 GGGCCCTGCTATTGCTAAAGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGCAAC 2997
          | || ||||| ||  |  |  | ||| || |||| | ||||||  | ||||||||  ||
Db   1370 GAGCTCTGCTGTTAATTGAAGACAAGATCAAGACAAGAGGAGTCTTAAGGCCTCTCGAAG 1429

Qy   2998 CGGAAATCTACGTTCCAGCATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGA 3057
          | ||  | ||  | ||||| ||||| || ||| |  | |  || || |||||| | ||||
Db   1430 CAGAGGTGTATTTGCCAGCTTTGGATATATTGCAAGCATATGGTATAAAGCTGATGGAGA 1489

Qy   3058 AAGTGGAGACTTGAAAGTTCCCTG 3081
          | |  ||    | ||||  | |||
Db   1490 AGGCAGAATGATCAAAGAACTCTG 1513


RESULT  14
AF271636
LOCUS       AF271636               15146 bp    DNA     linear   PLN 08-OCT-2000
DEFINITION  Zea mays lysine ketoglutarate reductase/saccharopine dehydrogenase
            (LKRSDH) gene, complete cds.
ACCESSION   AF271636
VERSION     AF271636.1  GI:10716964
KEYWORDS    .
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 15146)
  AUTHORS   Arruda,P., Kemper,E.L., Papes,F. and Leite,A.
  TITLE     Regulation of lysine catabolism in higher plants
  JOURNAL   Trends Plant Sci. 5 (8), 324-330 (2000)
  MEDLINE   20371120
   PUBMED   10908876
REFERENCE   2  (bases 1 to 15146)
  AUTHORS   Martinez-Moraes,K.C., Kemper,E.L., da Silva,F.R., Vettore,A.L. and
            Arruda,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2000) CBMEG, Unicamp, Campinas, SP 13083-970,
            Brazil
FEATURES             Location/Qualifiers
     source          1. .15146
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
     gene            4027. .15146
                     /gene="LKRSDH"
     CDS             join(4027. .4047,4150. .4413,4518. .4604,4697. .4826,
                     5080. .5253,5874. .5953,6202. .6318,6671. .6746,7396. .7567,
                     7875. .7971,8425. .8588,9379. .9433,9543. .9683,9870. .10190,
                     10390. .10470,10674. .10739,10861. .10978,11073. .11202,
                     11458. .11537,11868. .12026,12117. .12369,12561. .12669,
                     13060. .13221,13331. .13400,13486. .13541)
                     /gene="LKRSDH"
                     /note="bifunctional enzyme; LKR/SDH; lysine 2-oxoglutarate
                     reductase/saccharopine dehydrogenase"
                     /codon_start=1
                     /product="lysine ketoglutarate reductase/saccharopine
                     dehydrogenase"
                     /protein_id="AAG21985.1"
                     /db_xref="GI:10716965"
                     /translation="MGSAATEGNDTLLGNGVVGILAETCNMWERRAPLTPSHCARLLL
                     GGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQMILSD
                     RAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLID
                     FLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF
                     VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTS
                     RDMVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQ
                     LQQLMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVV
                     CLAVDILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLY
                     EYIPRMRNTMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHL
                     VRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIG
                     KVNEYETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVA
                     SLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIE
                     LKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKI
                     KAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYES
                     AKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATLRYEGFSEIMVTLSKT
                     GFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGC
                     CKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHH
                     EVEVEYPDGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKG
```

```
                    VIRPLQPEIYVPALEILESSGIKLVEKVET"
     3'UTR          13542..15146
                    /gene="LKRSDH"
BASE COUNT     4043 a   3133 c   3298 g   4672 t
ORIGIN

  Query Match             10.0%;  Score 326.6;  DB 8;  Length 15146;
  Best Local Similarity   96.0%;  Pred. No. 5.6e-73;
  Matches  335;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy     1439 GGATGATATGTCATACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAA 1498
            | |  |||||  | || |    ||  | ||||||||||||||||||||||||||||||||
Db     9842 GAAACATATGATACACACCTTTCTGCAGGTAGGAGCAGATGATACTGCCACATTGGATAA 9901

Qy     1499 AATTATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCA 1558
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9902 AATTATTGATTCCTTGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCA 9961

Qy     1559 AGAAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAAT 1618
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9962 AGAAATTGAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAAT 10021

Qy     1619 TGATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGC 1678
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10022 TGATAAAGGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGC 10081

Qy     1679 TGAGTTTCTGGCATCTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGA 1738
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10082 TGAGTTTCTGGCATCTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGA 10141

Qy     1739 TCAAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAG 1787
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db    10142 TCAAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAG 10190


RESULT  15
BD010450
LOCUS       BD010450                 619 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Chimeric genes and methods for increasing the lysine content of the
            seeds  of plants Chimeric genes and methods for increasing the
            lysine content of the seeds  of plants.
ACCESSION   BD010450
VERSION     BD010450.1  GI:18638823
KEYWORDS    JP 2001502923-A/82.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 619)
  AUTHORS   Falco,S.C., Iii,R.E.M. and Epelbaum,S.U.
  TITLE     Chimeric genes and methods for increasing the lysine content of the
            seeds of plants
  JOURNAL   Patent: JP 2001502923-A 82 06-MAR-2001;
            EI DU PONT DE NEMOURS AND CO
COMMENT     OS   Unidentified
            PN   JP 2001502923-A/82
            PD   06-MAR-2001
            PF   27-MAR-1998 JP 1998543284
            PR   27-MAR-1997 US    08/824627
            PI   SAVERIO CARL FALCO,RAYMOND ERVIN MCDEVITT III, PI   SABINE
            URSULA EPELBAUM
            PC   C12N9/06,C12N9/12,C12N9/88,C12P13/08,C12N15/82 CC
            Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1..619
            FT   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1..619
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      163 a    120 c    141 g    195 t
ORIGIN
```

```
  Query Match              9.2%;  Score 300.6;  DB 6;  Length 619;
  Best Local Similarity   69.8%;  Pred. No. 2e-66;
  Matches  435;  Conservative    0;  Mismatches  184;  Indels    4;  Gaps    2;

Qy        237 GAGAATATGCCACTGTTAGACAAGATCCTTGAAGAAAGGGTGTCCTTGTTTGATTATGAG 296
              || || |||||  || | || ||||| || |  || ||||  || || | ||| |||||
Db          1 GAAAACATGCCTTTGCTGGATAAGATTCTAGCTGAGAGGGCATCGTTATATGACTATGAA 60

Qy        297 CTAATTGTTGGAGATGATGGGAAAAGATCACTAGCATTTGGGAAATTTGCTGGTAGAGCT 356
               |||||||||| ||   ||||||||| | ||| |||||||| ||||| |||||||| |||
Db         61 TTAATTGTTGGGGACACTGGGAAAAGGTTACTTGCATTTGGAAAATTCGCTGGTAGGGCT 120

Qy        357 GGACTGATAGATTTCTTACATGGTCTCGGACAGCGATATTTGAGCCTTGGATACTCGACT 416
              ||| |||| || || || |  ||  | |||||||| | ||| || |||||||| || ||
Db        121 GGAATGATCGACTTTTTGCGCGGATTAGGACAGCGGTTTTTAAGTCTTGGATATTCAACA 180

Qy        417 CCATTTCTCTCTCTGGGACAATCTCATATGTATCCTTCGCTCGCTGCAGCCAAGGCTGCA 476
              || ||  | || || |||  |||| | ||||| ||||| || ||||| || ||||||||
Db        181 CCTTTCTTGTCACTTGGATCATCTTACATGTACCCTTCCCTGGCTGCTGCTAAGGCTGCT 240

Qy        477 GTCATTGTCGTTGCAGAAGAGATAGCAACATTTGGACTTCCATCCGGAATTTGTCCGATA 536
              || |||   ||||  ||| | || || ||    ||| | ||||  || ||||||||  |
Db        241 GTGATTTCTGTTGGTGAA-AAATTGCGACGCAGGGATTGCCATTGGGGATTTGTCCCCTG 299

Qy        537 GTGTTTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTCAAGTTA 596
              || | | | || |||||    ||||| || | |   ||||| ||||||||||| ||| |
Db        300 GTTTGTTTATTTACTGGTTCAGGAAATGTTTGTTCTGGTGCACAGGAGATATTTAAGCTT 359

Qy        597 TTGCCCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGGAATCTG 656
               | ||  ||||||||||||||| |    || || | |||  |    ||    | | | |
Db•       360 CTTCCTCATACCTTTGTTGATCCATCTAAACTACGCGACCTA---CATAGAACGGACCCA 416

Qy        657 TCTAAGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTGACCTCT 716
                | | | |   ||  |  | || ||||| || ||| |||||||||||||||||||  |
Db        417 GATCAACCAAGGCATGCTTCAAAAAGAGTTTTCCAAGTTTATGGTTGTGTTGTGACTGCC 476

Qy        717 AGAGACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTATTATGCT 776
                |||||| |||   | ||| |||| |   | |   ||||||||||  ||||| |||||
Db        477 CAAGACATGGTTGAACCCAAAGATCACGTGATAGTGTTTGACAAAGCAGACTACTATGCA 536

Qy        777 CATCCAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCTGTCATC 836
              ||||| || || ||||  ||   ||| ||||||| ||| || |||||||||||||| ||
Db        537 CATCCTGAGCATTACAATCCCACTTTCCATGAAAAAATAGCACCATATGCATCTGTTATT 596

Qy        837 GTAAACTGTATGTATTGGGAGAA 859
              || || || ||||||||||| ||
Db        597 GTCAATTGCATGTATTGGGAAAA 619


Search completed: June  1, 2002, 02:04:30
Job time: 9850 sec
```

```
                        GenCore version 4.5
              Copyright (c) 1993 - 2000  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         May 31, 2002, 22:48:59 ; Search time 2448.65 Seconds
                                          (without alignments)
                                          17996.675 Million cell updates/sec

Title:          US-09-049-304A-120
Perfect score:  3265
Sequence:       1 ATTGTGCCCGCCTTCTGCTA..........TCAAATAATGCCGATCAGTT 3265

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters:      27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      EST:*
                1:  em_estba:*
                2:  em_esthum:*
                3:  em_estin:*
                4:  em_estmu:*
                5:  em_estov:*
                6:  em_estpl:*
                7:  em_estro:*
                8:  em_htc:*
                9:  gb_est1:*
                10:  gb_est2:*
                11:  gb_htc:*
                12:  gb_gss:*
                13:  em_gss_hum:*
                14:  em_gss_inv:*
                15:  em_gss_pln:*
                16:  em_gss_vrt:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                  Description
-------------------------------------------------------------------------
    1   827.6    25.3    851 10  BG836723             BG836723 Zm08_04d1
    2   520.6    15.9    532  9  AW057000            AW057000 660007B05
    3   492.8    15.1    856 10  BI952486             BI952486 HVSMEm000
    4   490.4    15.0    525  9  AW681014            AW681014 WS1_8_B05
    5   433.2    13.3    635 10  BF200255             BF200255 WHE2254_G
    6   402.2    12.3    449 10  BF317962             BF317962 OV1_10_A0
    7   395.8    12.1    605 10  BG906544             BG906544 TaLr1150D
c   8     387    11.9    607 10  BG724449             BG724449 EST0006 D
    9   386.6    11.8    687 10  BI955770             BI955770 HVSMEm002
c  10     376    11.5    644  9  AV928505            AV928505 AV928505
   11   374.8    11.5    606  9  AV926619            AV926619 AV926619
   12     363    11.1    594  9  AV923326            AV923326 AV923326
   13   362.4    11.1    575 10  BG607870             BG607870 WHE2473_E
   14   355.2    10.9    613 10  BE606591             BE606591 WHE0903_C
   15   348.8    10.7    731 10  BI421952             BI421952 EST532618
   16   347.2    10.6    444 10  BG048770             BG048770 OV1_22_G0
   17   347.2    10.6    765 10  BM412142             BM412142 EST586469
```

```
   18   339.6    10.4    827 10  BF271904             BF271904 GA__Eb001
   19   335.6    10.3    527 10  BF259465             BF259465 HVSMEf001
   20   327.4    10.0    649  9  AI486200            AI486200 EST244521
   21   321.4     9.8    719 10  BM413549             BM413549 EST587876
   22   315.2     9.7    596 10  BG049063             BG049063 OV1_22_G0
   23   308.4     9.4    610  9  AI894899            AI894899 EST264342
   24   299.6     9.2    627  9  AI488742            AI488742 EST247081
   25     299     9.2    546 10  BE498116             BE498116 WHE0953_C
   26     295     9.0    685  9  AI054604            AI054604 coau0001I
   27   294.4     9.0    703 10  BF479283             BF479283 L48-2946T
   28   291.4     8.9    765 10  BI435118             BI435118 EST537879
   29   288.8     8.8    505  9  AL506473            AL506473 AL506473
   30   287.6     8.8    660 10  BG451844             BG451844 NF101B11D
   31   286.4     8.8    502 10  BG370012             BG370012 HVSMEi002
   32   284.8     8.7    642 10  BG890456             BG890456 EST516307
   33   280.6     8.6    395 10  BE428366             BE428366 MTD006.C0
   34   280.4     8.6    727 10  BI422689             BI422689 EST533355
   35   277.4     8.5    521  9  AW681093            AW681093 WS1_8_B05
   36   276.6     8.5    595 10  BM408787             BM408787 EST583114
   37   274.8     8.4    608  9  AW034280            AW034280 EST277851
   38   266.4     8.2    471 10  BM100629             BM100629 EBma01_SQ
   39   265.4     8.1    701 10  BM437056             BM437056 VVA013E08
   40   261.4     8.0    497 10  BE435871             BE435871 EST407045
   41     257     7.9    575  9  AI894874            AI894874 EST264317
   42   256.4     7.9    521 10  BE434479             BE434479 EST405557
   43   255.4     7.8    575  9  AW932187            AW932187 EST358030
   44   253.4     7.8    496  9  AW933092            AW933092 EST358935
   45   240.4     7.4    568 10  BI921398             BI921398 EST541301

                              ALIGNMENTS

RESULT   1
BG836723
LOCUS       BG836723                851 bp    mRNA    linear   EST 25-MAY-2001
DEFINITION  Zm08_04d10_A
            Zm08_AAFC_ECORC_Fusarium_graminearum_inoculated_corn_ear Zea mays
            cDNA clone Zm08_04d10, mRNA sequence.
ACCESSION   BG836723
VERSION     BG836723.1  GI:14203046
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 851)
  AUTHORS   Harris,L.J., Balcerzak,M., Allard,S., Saparno,A., Couroux,P., De
            Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S., Singh,J.A, Sprott
            ,D. and Tinker,N.A.
  TITLE     Expressed Sequence Tags from Developing Maize Kernels Six Days
            after Silk Channel Inoculation with Fusarium graminearum
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Harris, Linda J.
            Eastern Cereal and Oilseed Research Centre
            Agriculture and Agri-food Canada
            Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
            CANADA
            Tel: (613) 759-1314
            Fax: (613) 759-6566
            Email: harrislj@em.agr.ca.
FEATURES             Location/Qualifiers
     source          1..851
                     /organism="Zea mays"
                     /cultivar="CO430"
                     /db_xref="taxon:4577"
                     /clone="Zm08_04d10"
                     /clone_lib="Zm08_AAFC_ECORC_Fusarium_graminearum_inoculate
                     d_corn_ear"
                     /tissue_type="Developing kernels (sibcrossed)"
                     /dev_stage="10-11 days post-silk emergence"
                     /note="Vector: Bluescript SK+/XhoI-EcoRI; Site_1: EcoRI;
```

```
                     Site_2: XhoI; Field-grown maize ears were silk
                     channel-inoculated in the morning (~10 am) with 1 ml of a
                     Fusarium graminearum macroconidial suspension (500,000
                     spores/ml) and whole ears were collected and immediately
                     frozen in liquid nitrogen 6 days later."
BASE COUNT      250 a    175 c    204 g    213 t      9 others
ORIGIN

  Query Match             25.3%;  Score 827.6;  DB 10;  Length 851;
  Best Local Similarity   98.1%;  Pred. No. 4.6e-217;
  Matches  836;  Conservative   10;  Mismatches    5;  Indels    1;  Gaps    1;

Qy       2316 TTTGCTCTGGAACACTTGCCAAATCGGAATTCCTTGATATATGGTGACCTTTATGGTATC 2375
              |||||||||||| |||   | |||||||||||||||||||||||||||||||||||||||
Db          1 TTTGCTCTGGAAAACTGCCAAAATCGGAATTCCTTGATATATGGTGACCTTTATGGTATC 60

Qy       2376 TCCAAAGAAGCATCCACCATATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATT 2435
              |||||||||||||||||||||||||:|||||||:||||||||||||||||||||||||||
Db         61 TCCAAAGAAGCATCCACCATATATARGGCTACTCTTCGTTACGAAGGTTTTAGTGAGATT 120

Qy       2436 ATGGTAACCCTTTCCAAAACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGCAAGAT 2495
              |||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGGTAACCCTTTCCAAMACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGCAAGAT 180

Qy       2496 ACTAGTCGTCCAACATATAAGGGTTTCCTTGATGAACTACTGAATAATATCTCCACAATT 2555
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||:||||||
Db        181 ACTAGTCGTCCAACATATAAGGGTTTCCTTGATGMACTACTGAATAATATCTCMACAATT 240

Qy       2556 AACACGGACTTAGATATTGAAGCTTCTGGTGGATACGATGATGACCTGATTGCCAGACTG 2615
              ||:|||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        241 AASACGGACTTAGATATTGAAGCCTCTGGTGGATACGATGATGACCTGATTGCCAGACTG 300

Qy       2616 TTGAAGCTCGGGTGTTGCAAAAATAAGGAAATAGCTGTTAAGACAGTCAAAACCATCAAG 2675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGAAGCTCGGGTGTTGCAAAAATAAGGAAATAGCTGTTAAGACAGTCAAAACCATCAAG 360

Qy       2676 TTCTTGGGACTACATGAAGAGACTCAAATACCTAAGGGTTGTTCGAGCCCATTTGATGTG 2735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTCTTGGGACTACATGAAGAGACTCAAATACCTAAGGGTTGTTCGAGCCCATTTGATGTG 420

Qy       2736 ATTTGCCAGCGAATGGAACAGAGGATGGCCTATGGCCACAATGAGCAAGACATGGTACTG 2795
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        421 ATTTGCCAGCGAATGGAACAGAGGAT-GCCTATGGCCACAATGAGCAAGACATGGTACTG 479

Qy       2796 CTCCACCACGAAGTCGAGGTGGAATACCCGGACGGGCAACCCGCCGAAAAGCACCAAGCG 2855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 CTCCACCACGAAGTCGAGGTGGAATACCCGGACGGGCAACCCGCCGAAAAGCACCAAGCG 539

Qy       2856 ACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCACCACTGCCATGGCGCTGACC 2915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 ACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCACCACTGCCATGGCGCTGACC 599

Qy       2916 GTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAGGTCCAGACGAAA 2975
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 GTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAAAGAATAAGGTCCAGACGAAA 659

Qy       2976 GGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAGCATTGGAGATCTTGGAGTCG 3035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 GGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAGCATTGGAGATCTTGGAGTCG 719

Qy       3036 TCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGAAAGTTCCCTGATACACAGATAAAG 3095
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 TCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGAAAGTTCCCTGATACACAGATAAAG 779

Qy       3096 ATAGTATGATATAGCAGGGCACATGTATCTTTTGTATTAACTCCGTTCTGGAATATATAT 3155
              |||||||||||||||:|||:||||||:|:|||||||||||||||||||||||||||||||
Db        780 ATAGTATGATATAGCRGGGBACATGTRTSTTTTGTATTAACTCCGTTCTGGAATATATAT 839

Qy       3156 TTGTGAACTAAA 3167
              ||||||||||||
Db        840 TTGTGAACTAAA 851
```

```
RESULT   2
AW057000
LOCUS       AW057000                 532 bp    mRNA    linear   EST 29-SEP-1999
DEFINITION  660007B05.y1 660 - Mixed stages of anther and pollen Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW057000
VERSION     AW057000.1  GI:5932639
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 660007  row: B  column: 05.
FEATURES             Location/Qualifiers
     source          1..532
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="660 - Mixed stages of anther and pollen"
                     /tissue_type="whole premieotic anthers to pollen shed"
                     /dev_stage="premieotic anthers to pollen shed"
                     /lab_host="XLOLR"
                     /note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
                     Site_2: XhoI; Anther and pollen cDNA library.
                     Directionally sequenced with 5' end at the EcoRI site.
                     Created by Amie Franklin."
BASE COUNT      157 a    116 c    111 g    148 t
ORIGIN

  Query Match             15.9%;  Score 520.6;  DB 9;  Length 532;
  Best Local Similarity   99.8%;  Pred. No. 2.2e-132;
  Matches  520;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2046 GGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAAAGGGAAAA 2105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         12 GGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAAAGGGAAAA 71

Qy       2106 ATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAAACAATCCG 2165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 ATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAAACAATCCG 131

Qy       2166 CTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAAATCCTGCA 2225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 CTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAAATCCTGCA 191

Qy       2226 GTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATGAATCAGCA 2285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 GTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATGAATCAGCA 251

Qy       2286 AAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAAATCGGAAT 2345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 AAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAAATCGGAAT 311

Qy       2346 TCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATATATAGGGCT 2405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 TCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATATATAGGGCT 371

Qy       2406 ACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTGGGTTCTTT 2465
```

```
          |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    372 ACTCTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTGGGTTCTTT 431

Qy   2466 GATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATAAGGGTTTCCTT 2525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    432 GATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATAAGGGTTTCCTT 491

Qy   2526 GATGAACTACTGAATAATATCTCCACAATTAACACGGACTT 2566
          |||||||||||||||||||||||||||||||||||||||||
Db    492 GATGAACTACTGAATAATATCTCCACAATTAACACGGACTT 532


RESULT   3
BI952486
LOCUS       BI952486                 856 bp    mRNA    linear   EST 19-OCT-2001
DEFINITION  HVSMEm0006K11f Hordeum vulgare green seedling EST library
            HVcDNA0014 (Blumeria infected) Hordeum vulgare cDNA clone
            HVSMEm0006K11f, mRNA sequence.
ACCESSION   BI952486
VERSION     BI952486.1  GI:16296337
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 856)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
            Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons
            ,J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Blumeria infected Morex (compatible) seedling
            cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 435
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 11
            High quality sequence stop: 602.
FEATURES             Location/Qualifiers
     source          1..856
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEm0006K11f"
                     /clone_lib="Hordeum vulgare green seedling EST library
                     HVcDNA0014 (Blumeria infected)"
                     /tissue_type="green seedling leaf"
                     /lab_host="TJC121"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Morex (mla) plants were greenhouse grown in the R
                     Wise lab at Iowa State University, Ames, IA; 7 day old
                     green seedlings were infected with isolate 5874 of
                     Blumeria graminis f. sp. hordei, and leaves were harvested
                     24, 48 and 72 hr post-inoculation and snap frozen (Wise).
                     In the TJ Close lab at the University of California,
                     Riverside, total RNA was prepared from each sample pool,
                     equal quantities of all three RNA pools were combined,
                     poly(A) RNA was purified from the mixture, one primary
                     unamplified cDNA library was made, and 1 million pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids
                     (Chin). Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT      237 a    162 c    235 g    222 t
ORIGIN


  Query Match             15.1%;  Score 492.8;  DB 10;  Length 856;
  Best Local Similarity   79.5%;  Pred. No. 1.4e-124;
  Matches  620;  Conservative    1;  Mismatches  153;  Indels    6;  Gaps    3;

Qy   1740 CAAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGT 1799
          ||  | |||||| | ||||||||| ||||||||||||||||||||||||||||||| ||
Db      7 CAGGTACATGTTGTTGTGGCATCTCTGTATCAAAAAGATGCAGAAGAGACAGTTGACGGA 66

Qy   1800 ATTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATCTT 1859
          ||  | ||| |||| ||  | ||||| |||||| | |||| || ||| |||||  |||||
Db     67 ATAAAGAATGCAACAGCAGCTCAGCTCGATGTTTCAGATACTGAAAGTCTTTCGAATCTT 126

Qy   1860 GTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTGCA 1919
          ||||| |||||||| || ||| | || | |  | ||||||||||||||||||||||||||
Db    127 GTTTCACAGGTTGATGTCGTAGTCAGGTGGGGGGCTGCTAGTTTTCATGCTGCCATTGCA 186

Qy   1920 GGAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCATG 1979
           |||||||||||||| | ||||||||| |||| || ||||||||||||||||| ||||||
Db    187 AGAGTATGCATAGAGCTCAAGAAGCACTTGGTCACAGCAAGCTATGTTGATGATTCCATG 246

Qy   1980 TCAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCCTA 2039
          ||||| |||   ||||||||  |||  ||||||||||||||||| ||||||||||||||
Db    247 TCAAAGTTGGAGCAAGCTGCGCAAGGAGCAGGTGTAACTATACTCTGTGAAATGGGCCTT 306

Qy   2040 GATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAAAG 2099
          |||||||||||||||||| || |||| ||||| |||||||||||||| ||||| |  ||
Db    307 GATCCTGGCATAGATCACATGTTGTCGATGAAAATGATTGATGAAGCACATGCTCAGAAT 366

Qy   2100 GGAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAGCAAAC 2159
          |||||||||||||| |||||||||| |||||||||| | |||||||||||||| ||||||
Db    367 GGAAAAATAAAGGCGTTTACATCTTTCTGTGGTGGACTTCCATCTCCAGCTGCCGCAAAC 426

Qy   2160 AATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGAAAAAT 2219
          ||||| || |||||||| ||||||||||  ||||| |||||  ||||  | || | ||||
Db    427 AATCCACTGGCCTATAAGTTCAGTTGGAGTCCAGCAGGTGCCATCCGAGCTGGAAGAAAT 486

Qy   2220 CCTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATGAA 2279
          ||||| ||||||||||||||||| ||||  ||  |||| ||||||  ||| || |||||
Db    487 CCTGCTGTCTACAAATTTCTTGGAGAGATCATTAATGTGGATGGTAGTAAATTATATGAG 546

Qy   2280 TCAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAAAT 2339
          |||||||| ||||||||||||| |||||| ||||||||||||||||| ||||||||||||
Db    547 TCAGCAAAAAGGCTCAGACTACCAGAGCTGCCAGCTTTTGCTCTGGAGCACTTGCCAAAT 606

Qy   2340 CGGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATATAT 2399
          || ||||||||||| ||||| ||| | || || |||||||||||||||  |  |     |
Db    607 CGAAATTCCTTGATGTATGGAGACTTGTAGGGGATCTCCAAAGAAGCA--CTACTGGGTT 664

Qy   2400 AGGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTGGG 2459
          ||||| |||:||  ||   | || ||  | ||   |  |    ||   | |  || ||||
Db    665 AGGGCCACTCTTGTTTTGAAGGGATTAGGAGAACATGGGTATTCC---TGCGGAAGTGGG 721

Qy   2460 TTCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATAAGGGT 2519
           | |||  |||   | || ||||  || | ||||  |||| ||| ||||  |||  | ||
Db    722 GTTTTTTGTGCCTAAGATGATCCCTTGGT-CAAGGAACTAATCGCCCAATTTATGGGTGT 780


RESULT   4
AW681014
LOCUS       AW681014                 525 bp    mRNA    linear   EST 19-JUL-2000
```

```
DEFINITION  WS1_8_B05.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   AW681014
VERSION     AW681014.1  GI:7554719
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 517
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..525
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      155 a    102 c    122 g    146 t
ORIGIN

  Query Match             15.0%;  Score 490.4;  DB 9;  Length 525;
  Best Local Similarity   96.9%;  Pred. No. 4.9e-124;
  Matches  500;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy   1735 CAGATCAAATTCATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTG 1794
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     10 CAGATCAAATTCATGTTATTGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTG 69

Qy   1795 ATGGTATTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAG 1854
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db     70 ATGGTATTGAAAATACAACTGCTACCCAGCTTGATGTTTCTGATATTGGAAGCCTTTCAG 129

Qy   1855 ATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCA 1914
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    130 ATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTACTAGTTTTCATGCTGCCA 189

Qy   1915 TTGCAGGAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAAT 1974
          ||||| |||||||||||||| | |||||||||||||||||||||||||||||||||||||
Db    190 TTGCAAGAGTATGCATAGAGCTCAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAAT 249

Qy   1975 CCATGTCAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGG 2034
          ||||||||||||||||||||||||| |||| |||||||||||||||||||||||||||||
Db    250 CCATGTCAAACTTGAGCCAAGCTGCTAAAGGTGCAGGTGTAACTATACTTTGTGAAATGG 309

Qy   2035 GCCTAGATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCAC 2094
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    310 GCCTAGATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCTC 369

Qy   2095 GAAAGGGAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAG 2154
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    370 GAAAGGGAAAAATAAAGACATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGCAG 429
```

```
Qy   2155 CAAACAATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGGGA 2214
          ||||||| || ||||||||||||||||||||||||||||||||||| |||||| ||||||
Db    430 CAAACAACCCACTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCCCTCCGGGCAGGGA 489

Qy   2215 AAAATCCTGCAGTCTACAAATTTCTTGGTGAGACGA 2250
          |||||||||| ||||||||||||||||| |||||||
Db    490 AAAATCCTGCTGTCTACAAATTTCTTGGAGAGACGA 525


RESULT   5
BF200255
LOCUS       BF200255                 635 bp    mRNA    linear   EST 16-APR-2001
DEFINITION  WHE2254_G05_N10ZE Triticum monococcum vegetative apex cDNA library
            Triticum monococcum cDNA clone WHE2254_G05_N10, mRNA sequence.
ACCESSION   BF200255
VERSION     BF200255.1  GI:11114917
KEYWORDS    EST.
SOURCE      Triticum monococcum.
  ORGANISM  Triticum monococcum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Anderson,O.D., Chao,S., Dubcovsky,J., Echenique,V., Han,P.S., Hsia
            ,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L.,
            Stamove,B. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Vegetative apex cDNA library from Triticum monococcum
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            west Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene T3 primer.
FEATURES             Location/Qualifiers
     source          1..635
                     /organism="Triticum monococcum"
                     /cultivar="DV92"
                     /db_xref="taxon:4568"
                     /clone="WHE2254_G05_N10"
                     /clone_lib="Triticum monococcum vegetative apex cDNA
                     library"
                     /tissue_type="Vegetative shoot apex"
                     /dev_stage="Three weeks-old plants"
                     /lab_host="E. coli XLOLR"
                     /note="Vector: Lambda pBK-CMV (Lambda Zap Express),
                     excised phagemid; Site_1: EcoRI; Site_2: XhoI; The tissue,
                     total RNA, and poly(A) RNA were prepared, a cDNA library
                     was  made, and the cDNA clones were in vivo excised at the
                     University of  California, Davis (V. Echenique, B. Stamova
                     , J. Dubcovsky).  Plasmid  DNA preparations and DNA
                     sequencing were performed in the OD Anderson  lab (all
                     other authors)."
BASE COUNT      197 a    134 c    152 g    152 t
ORIGIN

  Query Match             13.3%;  Score 433.2;  DB 10;  Length 635;
  Best Local Similarity   82.2%;  Pred. No. 3.4e-108;
  Matches  511;  Conservative    1;  Mismatches  104;  Indels    6;  Gaps    1;

Qy   2259 GATGGTCATAACTTGTATGAATCAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTT 2318
          ||||||  ||| || || || |||||||||||||||||||||| ||||||||| ||||||
Db     14 GATGGTAGTAAATTATACGAGTCAGCAAAGAGGCTCAGACTACCAGAGCTTCCCGCTTTT 73

Qy   2319 GCTCTGGAACACTTGCCAAATCGGAATTCCTTGATATATGGTGACCTTTATGGTATCTCC 2378
          || |||||||||||||||||||| ||||||||||| ||||| ||||| ||||| ||||||
```

```
Db         74 GCCCTGGAACACTTGCCAAATCGAAATTCCTTGATGTATGGAGACCTGTATGGGATCTCC 133

Qy       2379 AAAGAAGCATCCACCATATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATG 2438
              ||||||||||| ||  |||||||| | |||:||||||| ||||| ||||||||||| |||
Db        134 AAAGAAGCATCTACTGTATATAGGTCCACTCTTCGTTATGAAGGATTTAGTGAGATCATG 193

Qy       2439 GTAACCCTTTCCAAAACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACT 2498
              |  | |||  | |||| |||||| |||||||||| ||||||||||||||| ||||| |||
Db        194 GCTATCCTGGCGAAAATTGGGTTTTTTGATGCTGAAAATCATCCACTGCTACAAGAAACT 253

Qy       2499 AGTCGTCCAACATATAAGGGTTTCCTTGATGAACTACTGAAT------AATATCTCCACA 2552
              | ||| |||||||||| |  ||| ||  | ||||| || |||      ||| | ||||||
Db        254 AATCGCCCAACATATAGGATTTTTCTCAACGAACTCCTTAATGACAACAATGTATCCACA 313

Qy       2553 ATTAACACGGACTTAGATATTGAAGCTTCTGGTGGATACGATGATGACCTGATTGCCAGA 2612
                 |||||  |  || ||  |||||   |||| ||| | |||||||| |||||| | |||
Db        314 TCCAACACAAAGGTAAATGGTGAAGAAACTGGAGGACATGATGATGAACTGATTTCAAGA 373

Qy       2613 CTGTTGAAGCTCGGGTGTTGCAAAAATAAGGAAATAGCTGTTAAGACAGTCAAAACCATC 2672
              ||  ||| ||| |||  ||||||| | |||||| ||||||| |||| | |||||||||||
Db        374 CTCATGATGCTTGGGCATTGCAAAGAAAAGGAACTAGCTGTCAAGATACTCAAAACCATC 433

Qy       2673 AAGTTCTTGGGACTACATGAAGAGACTCAAATACCTAAGGGTTGTTCGAGCCCATTTGAT 2732
              ||||||||||| || ||||| ||||| || || ||||||| |||||| ||  ||||   |
Db        434 AAGTTCTTGGGGCTGCATGAGGAGACACAGATTCCTAAGGATTGTTCAAGTGCATTCAGT 493

Qy       2733 GTGATTTGCCAGCGAATGGAACAGAGGATGGCCTATGGCCACAATGAGCAAGACATGGTA 2792
              || |||||||| |||||||||||||| ||||||||||||||||||||||| || ||||||
Db        494 GTTATTTGCCAACGAATGGAACAGAGAATGGCCTATGGCCACAATGAGCAGGATATGGTA 553

Qy       2793 CTGCTCCACCACGAAGTCGAGGTGGAATACCCGGACGGGCAACCCGCCGAAAAGCACCAA 2852
              ||||||||||| |||||||||||||| ||||| ||||||| |||||| || |||||||||
Db        554 CTGCTCCACCATGAAGTCGAGGTGGAGTACCCCGACGGGCGACCCGCGGAGAAGCACCAA 613

Qy       2853 GCGACGCTACTGGAGTTCGGGA 2874
              || ||||| ||||||||||| |
Db        614 GCTACGCTGCTGGAGTTCGGAA 635


RESULT   6
BF317962
LOCUS       BF317962                 449 bp    mRNA    linear   EST 21-NOV-2000
DEFINITION  OV1_10_A06.b1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BF317962
VERSION     BF317962.1  GI:11266499
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: ovaries of varying immature stages
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 414
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..449
                     /organism="Sorghum bicolor"
```

```
                     /db_xref="taxon:4558"
                     /clone_lib="Ovary 1 (OV1)"
                     /note="Organ: Mix of ovaries of varying immature stages
                     from 8-week-old plants; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      136 a     91 c     98 g    124 t
ORIGIN

  Query Match             12.3%;  Score 402.2;  DB 10;  Length 449;
  Best Local Similarity   93.3%;  Pred. No. 1e-99;
  Matches  419;  Conservative    1;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       2236 TTCTTGGTGAGACGATCCATGTAGATGGTCATAACTTGTATGAATCAGCAAAGAGGCTCA 2295
              ||||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db          1 TTCTTGGAGAGACGATCCATGTAGATGGTCATAACTTGTTTGAATCAGCAAAGAGGCTCA 60

Qy       2296 GACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAAATCGGAATTCCTTGATAT 2355
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTACCAGAGCTTCCAGCTTTTGCTCTGGAACACTTGCCAAATCGGAATTCCTTGATAT 120

Qy       2356 ATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCACCATATATAGGGCTACTYTTCGTT 2415
              |||| ||||||||||||||||||||||||||||| |||||||| |||||||||:||||||
Db        121 ATGGAGACCTTTATGGTATCTCCAAAGAAGCATCTACCATATACAGGGCTACTCTTCGTT 180

Qy       2416 ACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAAAACTGGGTTCTTTGATGCTGCAA 2475
              ||||||||||||||||||| |||| |||||| || |||| |||||| |||||||||||||
Db        181 ACGAAGGTTTTAGTGAGATAATGGCAACCCTGTCGAAAATTGGGTTGTTTGATGCTGCAA 240

Qy       2476 ATCATCCACTGCTGCAAGATACTAGTCGTCCAACATATAAGGGTTTCCTTGATGAACTAC 2535
              ||||||||||||||||||| |||| ||| |||||||||||||||||||||||||||||||
Db        241 ATCATCCACTGCTGCAAGAAACTAATCGCCCAACATATAAGGGTTTCCTTGATGAACTAC 300

Qy       2536 TGAATAATATCTCCACAATTAACACGGACTTAGATATTGAAGCTTCTGGTGGATACGATG 2595
              | |||||||||||||||| |||||| | ||||||||||||||| ||||||||||| ||||
Db        301 TTAATAATATCTCCACAACTAACACAGGCTTAGATATTGAAGCCTCTGGTGGATATGATG 360

Qy       2596 ATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTTGCAAAAATAAGGAAATAGCTGTTA 2655
              ||||  | ||||||||||||| ||||||||||||||||| | |||||| |||||||| ||
Db        361 ATGAAATCATTGCCAGACTGTCGAAGCTCGGGTGTTGCAGAGATAAGGGAATAGCTGCTA 420

Qy       2656 AGACAGTCAAAACCATCAAGTTCTTGGGA 2684
              |||||||||||||||||||||||||||||
Db        421 AGACAGTCAAAACCATCAAGTTCTTGGGA 449


RESULT   7
BG906544
LOCUS       BG906544                 605 bp    mRNA    linear   EST 05-JUN-2001
DEFINITION  TaLr1150D12R TaLr1 Triticum aestivum cDNA clone TaLr1150D12 5',
            mRNA sequence.
ACCESSION   BG906544
VERSION     BG906544.1  GI:14314220
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 605)
  AUTHORS   Cloutier,S., Dong,G. and Walsh,A.
  TITLE     Wheat functional genomics- Thatcher Lr1 cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
```

```
            Average inset size is >2.2 kb
            Plate: 150  row: D  column: 12
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..605
                     /organism="Triticum aestivum"
                     /cultivar="Thatcher Lr1"
                     /db_xref="taxon:4565"
                     /clone="TaLr1150D12"
                     /clone_lib="TaLr1"
                     /tissue_type="Leaf tissue"
                     /dev_stage="14 Days old"
                     /lab_host="E. coli XLOLR"
                     /note="Vector: Lambda ZapII; mass excised in plasmid
                     vector pBK-CMV (Stratagene).; Site_1: EcoRI; Site_2: XhoI;
                     mRNA obtained from wheat NIL Thatcher Lr1 24 hours after
                     inoculation  with leaf rust pathogen Puccinia triticina
                     race BBB carrying the  avirulence gene Avr1."
BASE COUNT      187 a    135 c    135 g    148 t
ORIGIN


  Query Match            12.1%;  Score 395.8;  DB 10;  Length 605;
  Best Local Similarity  81.1%;  Pred. No. 7.1e-98;
  Matches  473;  Conservative    1;  Mismatches  103;  Indels    6;  Gaps    1;

Qy   2131 GTGGATTGCCATCTCCAGCTGCAGCAAACAATCCGCTTGCCTATAAATTCAGTTGGAACC 2190
          ||||| | |||||||||||||| ||||||||||| || ||||| || ||||||||||  |
Db     23 GTGGACTTCCATCTCCAGCTGCCGCAAACAATCCACTGGCCTACAAGTTCAGTTGGAGTC 82

Qy   2191 CAGCTGGTGCACTCCGGTCAGGGAAAAATCCTGCAGTCTACAAATTTCTTGGTGAGACGA 2250
          |||| |||||  ||||  |||| | ||||||||| ||||||||||||||||| ||||| |
Db     83 CAGCAGGTGCCATCCGAGCAGGAAGAAATCCTGCTGTCTACAAATTTCTTGGAGAGACCA 142

Qy   2251 TCCATGTAGATGGTCATAACTTGTATGAATCAGCAAAGAGGCTCAGACTACGAGAGCTTC 2310
          || |||| ||||||  ||| || || || |||||||||||||||||||||| ||||||||
Db    143 TCAATGTGGATGGTAGTAAATTATACGAGTCAGCAAAGAGGCTCAGACTACCAGAGCTTC 202

Qy   2311 CAGCTTTTGCTCTGGAACACTTGCCAAATCGGAATTCCTTGATATATGGTGACCTTTATG 2370
          | |||||||| |||||||||||||||||||| ||||||||||| ||||| ||||| ||||
Db    203 CCGCTTTTGCCCTGGAACACTTGCCAAATCGAAATTCCTTGATGTATGGAGACCTGTATG 262

Qy   2371 GTATCTCCAAAGAAGCATCCACCATATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTG 2430
          | ||||||||||||||||| ||| |||||||| | |||:||||||| ||||| |||||||
Db    263 GGATCTCCAAAGAAGCATCTACCGTATATAGGTCCACTCTTCGTTATGAAGGATTTAGTG 322

Qy   2431 AGATTATGGTAACCCTTTCCAAAACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGC 2490
          |||| ||||  | |||  | |||| |||||| |||||||||| ||||||||||||||| |
Db    323 AGATCATGGCTATCCTGGCGAAAATTGGGTTTTTTGATGCTGAAAATCATCCACTGCTAC 382

Qy   2491 AAGATACTAGTCGTCCAACATATAAGGGTTTCCTTGATGAACTACTGAAT------AATA 2544
          |||| |||| ||| |||||||||| |  ||| ||  | ||||| || |||      |||
Db    383 AAGAAACTAATCGCCCAACATATAGGATTTTTCTCAACGAACTCCTTAATGTCAACAATG 442

Qy   2545 TCTCCACAATTAACACGGACTTAGATATTGAAGCTTCTGGTGGATACGATGATGACCTGA 2604
          | ||||||  ||||||  |  || ||  |||||   |||| ||| | || ||||| ||||
Db    443 TATCCACATCTAACACAAAGGTAAATGGTGAAGAAACTGGAGGACATGACGATGAACTGA 502

Qy   2605 TTGCCAGACTGTTGAAGCTCGGGTGTTGCAAAAATAAGGAAATAGCTGTTAAGACAGTCA 2664
          || | |||||  ||| ||| |||  ||||||| | |||||| ||||||| |||| | |||
Db    503 TTTCAAGACTCATGATGCTTGGGCATTGCAAAGAAAAGGAACTAGCTGTCAAGATACTCA 562

Qy   2665 AAACCATCAAGTTCTTGGGACTACATGAAGAGACTCAAATACC 2707
          ||||||||||||||||||| || ||||| ||||| || || ||
Db    563 AAACCATCAAGTTCTTGGGGCTGCATGAGGAGACACAGATTCC 605


RESULT 8
BG724449/c
LOCUS       BG724449                 607 bp    mRNA    linear   EST 08-MAY-2001
DEFINITION  EST0006 Dactylis leaf DDRT-cDNA Dactylis glomerata cDNA clone
            5s8-T3 similar to Lysine ketoglutarate reductase, mRNA sequence.
ACCESSION   BG724449
```

```
VERSION     BG724449.1  GI:14007845
KEYWORDS    EST.
SOURCE      orchard grass.
  ORGANISM  Dactylis glomerata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Poeae; Dactylis.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   Trejo Calzada,R. and O'Connell,M.A.
  TITLE     Drought induced transcripts in leaves of Dactylis glomerata
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mary A. O'Connell
            Department of Agronomy and Horticulture
            New Mexico State University
            MSC 3Q, P.O. Box 30003, Las Cruces, NM 88003, USA
            Tel: 505 646 5172
            Fax: 505 646 6041
            Email: moconnel@nmsu.edu
            Insert Length: 750   Std Error: 0.00
            Seq primer: T3
            High quality sequence stop: 607.
FEATURES             Location/Qualifiers
     source          1..607
                     /organism="Dactylis glomerata"
                     /db_xref="taxon:4509"
                     /clone="5s8-T3"
                     /clone_lib="Dactylis leaf DDRT-cDNA"
                     /dev_stage="drought-stressed"
                     /note="Organ: leaf; Vector: pGEM-T"
BASE COUNT      179 a    134 c    125 g    169 t
ORIGIN


  Query Match            11.9%;  Score 387;  DB 10;  Length 607;
  Best Local Similarity  78.8%;  Pred. No. 1.9e-95;
  Matches  475;  Conservative    0;  Mismatches  125;  Indels    3;  Gaps    1;

Qy    531 CCGATAGTGTTTGTGTTCACTGGAGTTGGAAACGTCTCTCAGGGTGCGCAGGAGATATTC 590
          || ||||| ||||| ||||| |||  |||||| || || |||||  | |||||||||||
Db    603 CCAATAGTATTTGTATTCACGGGATCTGGAAATGTTTCGCAGGGCACACAGGAGATATTT 544

Qy    591 AAGTTATTGCCCCATACCTTTGTTGATGCTGAGAAGCTTCCCGAAATTTTTCAGGCCAGG 650
          ||| |||||||||||||||| ||||||||||| || ||||| || |||| | ||| || |
Db    543 AAGCTATTGCCCCATACCTTCGTTGATGCTGACAAACTTCCTGAGATTTCTAAGGGCAAG 484

Qy    651 AATCTGTCTAAGCAATCTCAGTCGACCAAGAGAGTATTTCAACTTTATGGTTGTGTTGTG 710
          | || ||||   | |  | |||| |||| ||||| ||| ||||| ||||| ||||||||
Db    483 AGTCCGTCT---CCACATGAGTCAACCAGGAGAGCATTCCAACTATATGGATGTGTTGTC 427

Qy    711 ACCTCTAGAGACATAGTTTCTCACAAGGATCCCACCAGACAATTTGACAAAGGTGACTAT 770
          || || ||||| || | ||| | ||||||||||| || |   || |||||||| |||||||
Db    426 ACATCGAGAGATATGGTCTCACCCAAGGATCCCAGCAAATGTTTCGACAAAGATGACTAT 367

Qy    771 TATGCTCATCCAGAACACTACACCCCTGTTTTTCATGAAAGAATTGCTCCATATGCATCT 830
          |||||||| |||||||||||||| |||||||||| ||||||||| ||| ||||||||||||||
Db    366 TATGCTCACCCAGAACACTACAACCCTGTTTTCCATGAAAGGATTACTCCATATGCATCT 307

Qy    831 GTCATCGTAAACTGTATGTATTGGGAGAAGAGGTTTCCACCATTACTAAATATGGATCAG 890
          | ||| || || |||||||| ||||||| |||||||||||   || | |  |  || |||
Db    306 GCCATTGTTAATTGTATGTACTGGGAGAGGAGGTTTCCACGGCTATTGAGCACTGAACAG 247

Qy    891 TTACAGCAATTGATGGAGACTGGTTGTCCTTTAGTCGGCGTTTGTGACATAACTTGTGAT 950
          ||||| ||  ||| |   | ||| || ||||| || |||||||| || |||||||||||
Db    246 TTACAACAGCTGACGACAAATGGATGCCCTTTGGTTGGCGTTTGCGATATAACTTGTGAC 187

Qy    951 ATTGGAGGTTCCATTGAATTTATCAACAAGAGTACATCAATAGAGAGGCCTTTCTTTCGG 1010
          |||||||||||||| ||||||| | || ||||||||||| ||||||||||||||||||| |||
Db    186 ATTGGAGGTTCCATAGAATTTGTGAATAAGAGTACATCGATAGAGAGGCCTTTCTTCCGG 127

Qy   1011 TATGATCCTTCTAAGAATTCATACCATGATGATATGGAAGGTGCCGGAGTGGTCTGCTTG 1070
          || ||||||||||  |||| |||| ||||||||||||||||    || ||| | |||||
Db    126 TACGATCCTTCTACTAATTTATACTATGATGATATGGAAGGCAATGGTGTGATTTGCTTA 67
```

```
Qy    1071 GCTGTTGACATTCTCCCTACAGAATTCTCTAAAGAGGCCTCCCAACATTTTGGAAACATA 1130
           |||||||||||||| |||||||| || |||| |||| |  ||     ||   ||    ||
Db      66 GCTGTTGACATTCTACCTACAGAGTTTTCTAGAGAGCCTACCAGCATTTAGTGAGGGTTA 7

Qy    1131 CTA 1133
            ||
Db       6 ATA 4


RESULT   9
BI955770
LOCUS       BI955770                 687 bp    mRNA    linear   EST 19-OCT-2001
DEFINITION  HVSMEm0024G22f Hordeum vulgare green seedling EST library
            HVcDNA0014 (Blumeria infected) Hordeum vulgare cDNA clone
            HVSMEm0024G22f, mRNA sequence.
ACCESSION   BI955770
VERSION     BI955770.1  GI:16302375
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
            Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons
            ,J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Blumeria infected Morex (compatible) seedling
            cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 429
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 2
            High quality sequence stop: 654.
FEATURES             Location/Qualifiers
     source          1..687
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEm0024G22f"
                     /clone_lib="Hordeum vulgare green seedling EST library
                     HVcDNA0014 (Blumeria infected)"
                     /tissue_type="green seedling leaf"
                     /lab_host="TJC121"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Morex (mla) plants were greenhouse grown in the R
                     Wise lab at Iowa State University, Ames, IA; 7 day old
                     green seedlings were infected with isolate 5874 of
                     Blumeria graminis f. sp. hordei, and leaves were harvested
                     24, 48 and 72 hr post-inoculation and snap frozen (Wise).
                     In the TJ Close lab at the University of California,
                     Riverside, total RNA was prepared from each sample pool,
                     equal quantities of all three RNA pools were combined,
                     poly(A) RNA was purified from the mixture, one primary
                     unamplified cDNA library was made, and 1 million pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids
                     (Chin). Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT      196 a    141 c    169 g    180 t      1 others
ORIGIN


  Query Match             11.8%;  Score 386.6;  DB 10;  Length 687;
  Best Local Similarity   76.9%;  Pred. No. 2.6e-95;
  Matches  528;  Conservative    0;  Mismatches  119;  Indels   40;  Gaps    3;

Qy    1453 ACTCAGAGCTTGAAGTAGGAGCAGATGATACTGCCACATTGGATAAAATTATTGATTCCT 1512
           |||  ||      ||||||||||||||||||  | ||||| || || || ||||||||||
Db      11 ACTGCGATGAATCAGTAGGAGCAGATGATACGACTACATTAGACAAGATCATTGATTCCT 70

Qy    1513 TGACTTCTTTAGCTAATGAACATGGTGGAGATCACGATGCCGGGCAAGAAATTGAATTAG 1572
           |||||||| |||||||||  ||  |||||||||   ||||||           ||  ||
Db      71 TGACTTCTGTAGCTAATGCGCACCGTGGAGATCCTAATGCCG---------CCGAGATAT 121

Qy    1573 CTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAATTGATAAAGGAGGGC 1632
           |||| |||||||||| |||||  || |  | || ||| | ||   | | ||||| | | |
Db     122 CTCTAAAGATAGGAAGAGTCAGCGAATGCGGAATTGATGACAGCGTGGGTAAAGTACGAC 181

Qy    1633 CAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTGAGTTTCTGGCAT 1692
           |||||  |||||| ||||||||||||||||| |||||||| ||||||||||||||  |||
Db     182 CAAAGGGTTTAATCCTTGGAGCTGGAAGAGTTTGTCGGCCGGCTGCTGAGTTTCTAACAT 241

Qy    1693 CTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTA 1752
           |||||| | ||||                               || ||| | ||||||
Db     242 CTTACCAAAACAT------------------------------CGACCAAGTACATGTTG 271

Qy    1753 TCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAA 1812
           | ||||||||| ||||||||||||||||||||||||||||||| || ||  | ||| |||
Db     272 TTGTGGCATCTCTGTATCAAAAAGATGCAGAAGAGACAGTTGACGGAATAAAGAATGCAA 331

Qy    1813 CTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTG 1872
           | ||  | ||||| |||||| | |||| || ||| |||||  |||||||||| |||||||
Db     332 CAGCAGCTCAGCTCGATGTTTCAGATACTGAAAGTCTTTCGAATCTTGTTTCACAGGTTG 391

Qy    1873 AGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAGGAGTATGCATAG 1932
           | || ||| | |||||||||||||||||||||||||||||||||||| ||||||||||||
Db     392 ATGTCGTAGTCAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAAGAGTATGCATAG 451

Qy    1933 AGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCATGTCAAACTTGAGCC 1992
           || | |||||||||  ||| || ||||||||||||||||| ||||||||||| |||   |
Db     452 AGCTCAAGAAGCACTCGGTCACAGCAAGCTATGTTGATGATTCCATGTCAAAGTTGGAGC 511

Qy    1993 AAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCCTAGATCCTGGCATAG 2052
           |||||||  |||  ||||||||||||||||| |||||||||||||| |||||||||||||
Db     512 AAGCTGCGCAAGGAGCAGGTGTAACTATACTCTGTGAAATGGGCCTTGATCCTGGCATAG 571

Qy    2053 ATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAAAGGGAAAAATAAAGG 2112
           ||||| || |||| ||||| |||||||||||||| ||||| |  || || ||||||| ||
Db     572 ATCACATGTTGTCGATGAAAATGATTGATGAAGCACATGCTCCCAATGGGAAAATAAGGG 631

Qy    2113 CATTTACATCTTACTGTGGTGGATTGC 2139
           |  ||||||||| |||||||||| | |
Db     632 C-GTTACATCTTTCTGTGGTGGACTTC 657


RESULT  10
AV928505/c
LOCUS       AV928505                 644 bp    mRNA    linear   EST 18-JAN-2002
DEFINITION  AV928505 K. Sato unpublished cDNA library, cv. Haruna Nijo second
            leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA
            clone basdl01 3', mRNA sequence.
ACCESSION   AV928505
VERSION     AV928505.1  GI:18224302
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare.
```

```
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 644)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .644
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="basd101"
                     /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
                     Nijo second leaf stage seedling leaves"
                     /tissue_type="seedling leaves"
                     /dev_stage="second leaf stage"
BASE COUNT      130 a    181 c    141 g    192 t
ORIGIN


  Query Match            11.5%;  Score 376;  DB 9;  Length 644;
  Best Local Similarity  81.3%;  Pred. No. 2.1e-92;
  Matches  436;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy   2536 TGAATAATATCTCCACAATTAACACGGACTTAGATATTGAAGCTTCTGGTGGATACGATG 2595
          | || ||| | ||||||  ||||||  |  || ||  |||||   |||| ||| | ||||
Db    643 TCAACAATGTATCCACATCTAACACAAAGGTAAATGGTGAAGAAACTGGAGGACATGATG 584

Qy   2596 ATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTTGCAAAAATAAGGAAATAGCTGTTA 2655
          |||| |||||| | |||||  ||| ||| |||  ||||||| | |||||| ||||||| |
Db    583 ATGAACTGATTTCAAGACTAATGATGCTTGGGCATTGCAAAGAGAAGGAACTAGCTGTCA 524

Qy   2656 AGACAGTCAAAACCATCAAGTTCTTGGGACTACATGAAGAGACTCAAATACCTAAGGGTT 2715
          ||| | |||||||||| ||||||||||| || ||||| |||||  | || ||||||| ||
Db    523 AGATACTCAAAACCATAAAGTTCTTGGGGCTGCATGAGGAGACAGAGATTCCTAAGGATT 464

Qy   2716 GTTCGAGCCCATTTGATGTGATTTGCCAGCGAATGGAACAGAGGATGGCCTATGGCCACA 2775
          |||| ||  ||||   ||| |||||||| |||||||||||||| ||||||||||||||||
Db    463 GTTCAAGTGCATTCAGTGTTATTTGCCAACGAATGGAACAGAGAATGGCCTATGGCCACA 404

Qy   2776 ATGAGCAAGACATGGTACTGCTCCACCACGAAGTCGAGGTGGAATACCCGGACGGGCAAC 2835
          ||||||| || ||||||||||||||||| ||||| |||||||| ||||| ||||||| ||
Db    403 ATGAGCAGGATATGGTACTGCTCCACCATGAAGTGGAGGTGGAGTACCCCGACGGGCGAC 344

Qy   2836 CCGCCGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGAAGGTTGAAAATGGCAGGTCCA 2895
          || |||| ||||||||||| ||||| ||||||||||| |||   || || |||||||| |
Db    343 CCACCGAGAAGCACCAAGCAACGCTGCTGGAGTTCGGAAAGACCGAGAACGGCAGGTCAA 284

Qy   2896 CCACTGCCATGGCGCTGACCGTCGGCATTCCAGCAGCAATAGGGGCCCTGCTATTGCTAA 2955
          |||| |||||||| || ||||||||  | || ||||| ||||| |||||||| |||||
Db    283 CCACCGCCATGGCCCTCACCGTCGGGGTACCGGCAGCGATAGGAGCCCTGCTCTTGCTCC 224

Qy   2956 AGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGCAACCGGAAATCTACGTTCCAG 3015
          |||| ||||| |||| |||||| |||||| |||||||||||||||| |||||| | || |
Db    223 AGAACAAGGTTCAGAGGAAAGGCGTGATCCGGCCTCTGCAACCGGAGATCTACATCCCTG 164

Qy   3016 CATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGAAAGTGGAGACTTGA 3071
          |  |||||||||||||  ||||||||||||||||| | |||| |||| |||| |||
Db    163 CGCTGGAGATCTTGGATGCGTCGGGCATCAAGCTGATCGAGAGAGTGCAGACCTGA 108


RESULT  11
AV926619
LOCUS       AV926619                 606 bp    mRNA    linear   EST 18-JAN-2002
```

```
DEFINITION  AV926619 K. Sato unpublished cDNA library, cv. Haruna Nijo second
            leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA
            clone basd23b07 5', mRNA sequence.
ACCESSION   AV926619
VERSION     AV926619.1  GI:18222416
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare.
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="basd23b07"
                     /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
                     Nijo second leaf stage seedling leaves"
                     /tissue_type="seedling leaves"
                     /dev_stage="second leaf stage"
BASE COUNT      177 a    117 c    150 g    161 t      1 others
ORIGIN


  Query Match            11.5%;  Score 374.8;  DB 9;  Length 606;
  Best Local Similarity  78.3%;  Pred. No. 4.5e-92;
  Matches  492;  Conservative    0;  Mismatches   97;  Indels   39;  Gaps    2;

Qy   1467 GTAGGAGCAGATGATACTGCCACATTGGATAAAATTATTGATTCCTTGACTTCTTTAGCT 1526
          |||||||||||||||||  | ||||| || || || |||||||||||||||||| |||||
Db     17 GTAGGAGCAGATGATACGACTACATTAGACAAGATCATTGATTCCTTGACTTCTGTAGCT 76

Qy   1527 AATGAACATGGTGGAGATCACGATGCCGGGCAAGAAATTGAATTAGCTCTGAAGATAGGA 1586
          ||||  ||  |||||||||   ||||||           ||  || |||| |||||||||
Db     77 AATGCGCACCGTGGAGATCCTAATGCCG---------CCGAGATATCTCTAAAGATAGGA 127

Qy   1587 AAAGTCAATGAGTATGAAACTGACGTCACAATTGATAAAGGAGGGCCAAAGATTTTAATT 1646
          | |||||  || | | || ||| | ||  || ||||||| ||| |||||| |||||||
Db    128 AGAGTCAGCGAATGCGGAATTGATGACAGCATGGATAAAGTAGGACCAAAGGTTTTAATC 187

Qy   1647 CTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTGAGTTTCTGGCATCTTACCCAGACATA 1706
          ||||||||||||||||| |||||||| ||||||||||||||  ||||||||| | ||||
Db    188 CTTGGAGCTGGAAGAGTTTGTCGGCCGGCTGCTGAGTTTCTAACATCTTACCAAAACAT- 246

Qy   1707 TGTACCTATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATCTTTG 1766
                                        || ||| | |||||| | ||||||||| ||
Db    247 -----------------------------CGACCAAGTACATGTTGTTGTGGCATCTCTG 277

Qy   1767 TATCAAAAAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCAGCTT 1826
          ||||||||||||||||||||||||||||| || ||   | || |||| ||  | |||||
Db    278 TATCAAAAAGATGCAGAAGAGACAGTTGACGGAATAAAGAATGCAACAGCAGCTCAGCTC 337

Qy   1827 GATGTTGCTGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATTAGC 1886
          |||||| | |||| || ||| |||||  |||||||||| |||||||| || ||| | |||
Db    338 GATGTTTCAGATACTGAAAGTCTTTCGAATCTTGTTTCACAGGTTGATGTCGTAGTCAGC 397

Qy   1887 TTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAGGAGTATGCATAGAGTTGAAGAAGCAC 1946
          ||||||||||||||||||||||||||||||||| |||||||||||||| | |||||||||
Db    398 TTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAAGAGTATGCATAGAGCTCAAGAAGCAC 457

Qy   1947 ATGGTAACGGCAAGCTATGTTGATGAATCCATGTCAAACTTGAGCCAAGCTGCCAAAGAT 2006
```

```
           |||| || ||||||||||||||||| ||||||||||| |||   |||||||| |||
Db     458 TTGGTCACAGCAAGCTATGTTGATGATTCCATGTCAAAGTTGGAGCAAGCTGCGCAAGGA 517

Qy    2007 GCAGGTGTAACTATACTTTGTGAAATGGGCCTAGATCCTGGCATAGATCACTTGATGTCA 2066
           ||||||||||||||||| |||||||||||||| |||||||||||||||||| || ||||
Db     518 GCAGGTGTAACTATACTCTGTGAAATGGGCCTTGATCCTGGCATAGATCACATGTTGTCG 577

Qy    2067 ATGAAGATGATTGATGAAGCTCATGCAC 2094
           ||||| |||||||||||||| ||||| |
Db     578 ATGAAAATGATTGATGAAGCACATGCTC 605


RESULT  12
AV923326
LOCUS       AV923326                 594 bp    mRNA    linear   EST 18-JAN-2002
DEFINITION  AV923326 K. Sato unpublished cDNA library, cv. Haruna Nijo second
            leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA
            clone basd12101 5', mRNA sequence.
ACCESSION   AV923326
VERSION     AV923326.1  GI:18219105
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare.
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="basd12101"
                     /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
                     Nijo second leaf stage seedling leaves"
                     /tissue_type="seedling leaves"
                     /dev_stage="second leaf stage"
BASE COUNT      173 a    114 c    149 g    157 t      1 others
ORIGIN


  Query Match             11.1%;  Score 363;  DB 9;  Length 594;
  Best Local Similarity   79.8%;  Pred. No. 8e-89;
  Matches  454;  Conservative    0;  Mismatches   85;  Indels   30;  Gaps    1;

Qy    1566 GAATTAGCTCTGAAGATAGGAAAAGTCAATGAGTATGAAACTGACGTCACAATTGATAAA 1625
           ||  || |||| |||||||||| |||||  || |  | || ||| | ||  || ||||||
Db      56 GAGATATCTCTAAAGATAGGAAGAGTCAGCGAATGCGGAATTGATGACAGCATGGATAAA 115

Qy    1626 GGAGGGCCAAAGATTTTAATTCTTGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTGAGTTT 1685
           | ||| |||||| ||||||| ||||||||||||||||| |||||||| ||||||||||||
Db     116 GTAGGACCAAAGGTTTTAATCCTTGGAGCTGGAAGAGTTTGTCGGCCGGCTGCTGAGTTT 175

Qy    1686 CTGGCATCTTACCCAGACATATGTACCTATGGTGTTGATGACCATGATGCAGATCAAATT 1745
           ||  ||||||||| | ||||                               || ||| |
Db     176 CTAACATCTTACCAAAACAT------------------------------CGACCAAGTA 205

Qy    1746 CATGTTATCGTGGCATCTTTGTATCAAAAAGATGCAGAAGAGACAGTTGATGGTATTGAA 1805
           |||||| | ||||||||| ||||||||||||||||||||||||||||||| || ||  |
Db     206 CATGTTGTTGTGGCATCTCTGTATCAAAAAGATGCAGAAGAGACAGTTGACGGAATAAAG 265

Qy    1806 AATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTCAGATCTTGTTTCT 1865
           ||| |||| ||  | ||||| |||||| | |||| || ||| |||||  ||||||||||
Db     266 AATGCAACAGCAGCTCAGCTCGATGTTTCAGATACTGAAAGTCTTTCGAATCTTGTTTCA 325

Qy    1866 CAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAGGAGTA 1925
           |||||||| || ||| | |||||||||||||||||||||||||||||||||||| |||||
Db     326 CAGGTTGATGTCGTAGTCAGCTTGCTGCCTGCTAGTTTTCATGCTGCCATTGCAAGAGTA 385

Qy    1926 TGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGAATCCATGTCAAAC 1985
           ||||||||| | ||||||||| |||| || ||||||||||||||||| |||||||||||
Db     386 TGCATAGAGCTCAAGAAGCACTTGGTCACAGCAAGCTATGTTGATGATTCCATGTCAAAG 445

Qy    1986 TTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAATGGGCCTAGATCCT 2045
           |||   ||||||||  |||  ||||||||||||||||| |||||||||||||| ||||||
Db     446 TTGGAGCAAGCTGCGCAAGGAGCAGGTGTAACTATACTCTGTGAAATGGGCCTTGATCCT 505

Qy    2046 GGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGCACGAAAGGGAAAA 2105
           |||||||||||| || |||| ||||| |||||||||||||| ||||| |  || ||||||
Db     506 GGCATAGATCACATGTTGTCGATGAAAATGATTGATGAAGCACATGCTCAGAATGGAAAA 565

Qy    2106 ATAAAGGCATTTACATCTTACTGTGGTGG 2134
           |||||||| |||||||||| |||||||||
Db     566 ATAAAGGCGTTTACATCTTTCTGTGGTGG 594


RESULT  13
BG607870
LOCUS       BG607870                 575 bp    mRNA    linear   EST 17-APR-2001
DEFINITION  WHE2473_E07_J13ZS Triticum monococcum early reproductive apex cDNA
            library Triticum monococcum cDNA clone WHE2473_E07_J13, mRNA
            sequence.
ACCESSION   BG607870
VERSION     BG607870.1  GI:13657853
KEYWORDS    EST.
SOURCE      Triticum monococcum.
  ORGANISM  Triticum monococcum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   Anderson,O.D., Chao,S., Dubcovsky,J., Echenique,V., Han,P.S., Hsia
            ,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L.,
            Stamova,B. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Early reproductive apex cDNA library from Triticum
            monococcum
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .575
                     /organism="Triticum monococcum"
                     /cultivar="DV92"
                     /db_xref="taxon:4568"
                     /clone="WHE2473_E07_J13"
                     /clone_lib="Triticum monococcum early reproductive apex
                     cDNA library"
                     /tissue_type="Early reproductive apex"
                     /dev_stage="Seven week-old plants"
                     /lab_host="E. coli XLOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; The tissue, total RNA, and
                     poly(A) RNA were prepared from apex at  double-ridge stage
                     to terminal-spikelet stage during transition from
                     vegetative state to flower state, a cDNA library was made,
                     and the  cDNA clones were in vivo excised at the
                     University of California, Davis  (V. Echenique, B. Stamova
```

```
            , J. Dubcovsky).  Plasmid DNA preparations and  DNA
            sequencing were performed in the OD Anderson lab (all
            other authors)."
BASE COUNT      173 a    111 c    139 g    152 t
ORIGIN

  Query Match            11.1%;  Score 362.4;  DB 10;  Length 575;
  Best Local Similarity   78.0%;  Pred. No. 1.2e-88;
  Matches  479;  Conservative    0;  Mismatches   96;  Indels   39;  Gaps    2;

Qy   1469 AGGAGCAGATGATACTGCCACATTGGATAAAATTATTGATTCCTTGACTTCTTTAGCTAA 1528
          |||||||||||||||  | ||||| || || || || ||||||||||||||| | |||||
Db      1 AGGAGCAGATGATACGACTACATTAGACAAGATCATCGATTCCTTGACTTCTATCGCTAA 60

Qy   1529 TGAACATGGTGGAGATCACGATGCCGGGCAAGAAATTGAATTAGCTCTGAAGATAGGAAA 1588
          || |||  |||||||||   |||||         |  ||  || |||| ||||||||||
Db     61 TGCACACCGTGGAGATCCTAATGCC---------ACCGAGATATCTCTAAAGATAGGAAG 111

Qy   1589 AGTCAATGAGTATGAAACTGACGTCACAATTGATAAAGGAGGGCCAAAGATTTTAATTCT 1648
          |||||  || | || || ||| | ||  || ||||||| ||| |||||| ||||||| ||
Db    112 AGTCAGCGAATGTGGAATTGATGACAGCATGGATAAAGTAGGACCAAAGGTTTTAATCCT 171

Qy   1649 TGGAGCTGGAAGAGTCTGTCGGCCAGCTGCTGAGTTTCTGGCATCTTACCCAGACATATG 1708
          ||||||||||||||| |||||||||||||||||||||||  ||||||||| | ||||
Db    172 TGGAGCTGGAAGAGTTTGTCGGCCAGCTGCTGAGTTTCTAACATCTTACCAAAACAT--- 228

Qy   1709 TACCTATGGTGTTGATGACCATGATGCAGATCAAATTCATGTTATCGTGGCATCTTTGTA 1768
                                              || ||| | |||||| | ||||||||| ||||
Db    229 ---------------------------CGACCAAGTACATGTTGTTGTGGCATCTCTGTA 261

Qy   1769 TCAAAAAGATGCAGAAGAGACAGTTGATGGTATTGAAAATACAACTGCTACCCAGCTTGA 1828
          ||||||||||||||||||||||||||| || ||  | ||| |||| ||  | ||||| ||
Db    262 TCAAAAAGATGCAGAAGAGACAGTTGACGGAATAAAGAATGCAACAGCAGCTCAGCTCGA 321

Qy   1829 TGTTGCTGATATTGGAAGCCTTTCAGATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTT 1888
          |||| | |||| || ||| |||||  |||||||||| |||||||| |||||| | |||||
Db    322 TGTTTCAGATACTGAAAGTCTTTCGAATCTTGTTTCACAGGTTGATGTTGTAGTCAGCTT 381

Qy   1889 GCTGCCTGCTAGTTTTCATGCTGCCATTGCAGGAGTATGCATAGAGTTGAAGAAGCACAT 1948
          ||| ||||| ||||||||||||||||||||| |||||||||||||| | ||||||||| |
Db    382 GCTACCTGCCAGTTTTCATGCTGCCATTGCAAGAGTATGCATAGAGCTCAAGAAGCACTT 441

Qy   1949 GGTAACGGCAAGCTATGTTGATGAATCCATGTCAAACTTGAGCCAAGCTGCCAAAGATGC 2008
          ||| || ||||||||||||||||| ||||||||||| |||   ||||||||  |||  ||
Db    442 GGTCACAGCAAGCTATGTTGATGATTCCATGTCAAAGTTGGAGCAAGCTGCGCAAGGAGC 501

Qy   2009 AGGTGTAACTATACTTTGTGAAATGGGCCTAGATCCTGGCATAGATCACTTGATGTCAAT 2068
          ||||||||||||||| || ||||||||||| |||||||||||||||||| || |||| ||
Db    502 AGGTGTAACTATACTCTGCGAAATGGGCCTTGATCCTGGCATAGATCACATGTTGTCGAT 561

Qy   2069 GAAGATGATTGATG 2082
          ||| ||||||||||
Db    562 GAAAATGATTGATG 575


RESULT 14
BE606591
LOCUS       BE606591                613 bp    mRNA    linear   EST 22-AUG-2000
DEFINITION  WHE0903_C03_F05ZS Wheat 5-15 DAP spike cDNA library Triticum
            aestivum cDNA clone WHE0903_C03_F05, mRNA sequence.
ACCESSION   BE606591
VERSION     BE606591.1  GI:9883755
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - 5-15 DAP spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE0903_C03_F05"
                     /clone_lib="Wheat 5-15 DAP spike cDNA library"
                     /tissue_type="Spike"
                     /dev_stage="Adult plant"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                     greenhouse. Spikes at 5, 10 and 15 DAP were  harvested,
                     total RNA and poly(A) RNA were prepared, a cDNA library
                     was  made, and the cDNA clones were in vivo excised to
                     give pBluescript  phagemids in the TJ Close lab (Choi,
                     Close, Fenton) at the University  of California,
                     Riverside. Plasmid DNA preparations and DNA sequencing
                     were performed in the OD Anderson lab (all other authors
                     )."
BASE COUNT      178 a    135 c    178 g    122 t
ORIGIN

  Query Match            10.9%;  Score 355.2;  DB 10;  Length 613;
  Best Local Similarity   83.0%;  Pred. No. 1.2e-86;
  Matches  405;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy   2587 GATACGATGATGACCTGATTGCCAGACTGTTGAAGCTCGGGTGTTGCAAAAATAAGGAAA 2646
          || | || ||||| |||||| | | |||  ||| ||| |||  ||||||| | ||||||
Db      8 GACATGACGATGAACTGATTTCAAAACTCATGATGCTTGGGCATTGCAAAGAAAAGGAAC 67

Qy   2647 TAGCTGTTAAGACAGTCAAAACCATCAAGTTCTTGGGACTACATGAAGAGACTCAAATAC 2706
          ||||||| |||| | |||||||||||||||||||||| || ||||| ||||| || || |
Db     68 TAGCTGTCAAGATACTCAAAACCATCAAGTTCTTGGGGCTGCATGAGGAGACACAGATTC 127

Qy   2707 CTAAGGGTTGTTCGAGCCCATTTGATGTGATTTGCCAGCGAATGGAACAGAGGATGGCCT 2766
          |||||| |||||| ||  ||||   ||| |||||||| || ||||||||||| |||||||
Db    128 CTAAGGATTGTTCAAGTGCATTCAGTGTTATTTGCCAACGTATGGAACAGAGAATGGCCT 187

Qy   2767 ATGGCCACAATGAGCAAGACATGGTACTGCTCCACCACGAAGTCGAGGTGGAATACCCGG 2826
          |||||||||||||||| || ||||||||||||||||| |||||||||||||| ||||| |
Db    188 ATGGCCACAATGAGCAGGATATGGTACTGCTCCACCATGAAGTCGAGGTGGAGTACCCCG 247

Qy   2827 ACGGGCAACCCGCCGAAAAGCACCAAGCGACGCTACTGGAGTTCGGGAAGGTTGAAAATG 2886
          |||||| |||  |||||||||||||||||||||| ||||||||||| |||   || ||||
Db    248 ACGGGCGACCAACCGAAAAGCACCAAGCGACGCTGCTGGAGTTCGGAAAGACCGAGAATG 307

Qy   2887 GCAGGTCCACCACTGCCATGGCGCTGACCGTCGGCATTCCAGCAGCAATAGGGGCCCTGC 2946
          ||||||| ||||| |||||||| || ||||||||  | || ||||| ||||| |||||||
Db    308 GCAGGTCGACCACCGCCATGGCCCTCACCGTCGGGGTACCGGCAGCGATAGGAGCCCTGC 367

Qy   2947 TATTGCTAAAGAATAAGGTCCAGACGAAAGGAGTGATCAGGCCTCTGCAACCGGAAATCT 3006
          | |||||  |||| |||||||||| |||||| |||||| |||||||| | ||||| ||||
Db    368 TCTTGCTCCAGAACAAGGTCCAGAGGAAAGGGGTGATCCGGCCTCTGGAGCCGGAGATCT 427

Qy   3007 ACGTTCCAGCATTGGAGATCTTGGAGTCGTCGGGCATCAAGCTGGTTGAGAAAGTGGAGA 3066
          || | || ||  |||||||||||||  || |||||||||||||| | |||| ||||||||
Db    428 ACATCCCTGCGCTGGAGATCTTGGAAGCGGCGGGCATCAAGCTGATCGAGAGAGTGGAGA 487
```

```
Qy   3067 CTTGAAAG 3074
          | ||| ||
Db    488 CCTGAGAG 495


RESULT 15
BI421952
LOCUS       BI421952                 731 bp    mRNA    linear   EST 16-AUG-2001
DEFINITION  EST532618 tomato callus, TAMU Lycopersicon esculentum cDNA clone
            cLEC68F1 5' end, mRNA sequence.
ACCESSION   BI421952
VERSION     BI421952.1  GI:15195940
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 731)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.
            , Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
            ,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato callus tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1..731
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEC68F1"
                     /clone_lib="tomato callus, TAMU"
                     /tissue_type="callus"
                     /dev_stage="25-40 days old"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; supplier: Giovannoni laboratory; cLEC - Cotyledons
                     of seedlings 7-10 days post-germination were excised, cut
                     at both ends and placed on MS medium with no selection.
                     Mixed callus was harvested at 25 and 40 days and included
                     undifferentiated masses. Tomato Callus EST Library"
BASE COUNT      207 a    155 c    167 g    202 t
ORIGIN


  Query Match             10.7%;  Score 348.8;  DB 10;  Length 731;
  Best Local Similarity   67.6%;  Pred. No. 7.5e-85;
  Matches  488;  Conservative    1;  Mismatches  233;  Indels    0;  Gaps    0;

Qy   1793 TGATGGTATTGAAAATACAACTGCTACCCAGCTTGATGTTGCTGATATTGGAAGCCTTTC 1852
          | | | ||||  |||| |||  || |  ||||||||  || |   |   |     |||||
Db      5 TAAAGTTATTCCAAATGCAAAAGCAATTCAGCTTGACATTACGAGTCACGAGTCTCTTTC 64

Qy   1853 AGATCTTGTTTCTCAGGTTGAGGTTGTAATTAGCTTGCTGCCTGCTAGTTTTCATGCTGC 1912
                  |  |  ||||||| ||||| || ||||| |||||| |||| |  |||| ||
Db     65 TAGCTGGATCGCAGAGGTTGATGTTGTCATCAGCTTACTGCCTCCTAGCTGCCATGGTGT 124

Qy   1913 CATTGCAGGAGTATGCATAGAGTTGAAGAAGCACATGGTAACGGCAAGCTATGTTGATGA 1972
           || |||   | |||||| ||| ||||||| ||  | || ||||| ||||| ||||||||
Db    125 TATAGCAAAGGCATGCATTGAGCTGAAGAAACATCTCGTCACGGCTAGCTACGTTGATGA 184

Qy   1973 ATCCATGTCAAACTTGAGCCAAGCTGCCAAAGATGCAGGTGTAACTATACTTTGTGAAAT 2032
           || |||| |||  |   ||||| ||| |||  ||| ||  | ||||| ||| |||||||
Db    185 TTCTATGTTAAAGCTAGACCAAGATGCAAAATCTGCTGGAATTACTATTCTTGGTGAAAT 244

Qy   2033 GGGCCTAGATCCTGGCATAGATCACTTGATGTCAATGAAGATGATTGATGAAGCTCATGC 2092
          |||| | || || || ||||||||  ||||| | ||||| |||||  | ||||| |||||
Db    245 GGGCTTGGACCCAGGAATAGATCATATGATGGCGATGAAAATGATCAACGAAGCCCATGC 304

Qy   2093 ACGAAAGGGAAAAATAAAGGCATTTACATCTTACTGTGGTGGATTGCCATCTCCAGCTGC 2152
          |  ||||||||| || | | | ||    ||||||||||||||  | || |||||||||||
Db    305 AGCAAAGGGAAAGATCAGGTCTTTCGTTTCTTACTGTGGTGGTCTTCCCTCTCCAGCTGC 364

Qy   2153 AGCAAACAATCCGCTTGCCTATAAATTCAGTTGGAACCCAGCTGGTGCACTCCGGTCAGG 2212
           || ||||||||  | || ||||| ||||||||||  ||||| || ||  | ||  | ||
Db    365 TGCCAACAATCCATTAGCTTATAAGTTCAGTTGGAGTCCAGCCGGAGCTATACGAGCTGG 424

Qy   2213 GAAAAATCCTGCAGTCTACAAATTTCTTGGTGAGACGATCCATGTAGATGGTCATAACTT 2272
          |   ||||| |||| |||||  | ||  || || |  |||||||| || ||||| ||| |
Db    425 GTGGAATCCAGCAGCCTACAGGTATCAAGGGGAAATTATCCATGTCGAAGGTCAGAACCT 484

Qy   2273 GTATGAATCAGCAAAGAGGCTCAGACTACGAGAGCTTCCAGCTTTTGCTCTGGAACACTT 2332
           ||||| |||||    | |||  | || |  ||  ||||||| |||||  | ||     |
Db    485 TTATGATTCAGCTGCAAAGCTTCGTCTTCCTGATTTTCCAGCCTTTGCATTAGAGTGTAT 544

Qy   2333 GCCAAATCGGAATTCCTTGATATATGGTGACCTTTATGGTATCTCCAAAGAAGCATCCAC 2392
           |||||||| || |||||  | ||||| ||| | |||||||| ||  |||||||||| ||
Db    545 ACCAAATCGCAACTCCTTAGTGTATGGAGACTTGTATGGTATATCAGAAGAAGCATCTAC 604

Qy   2393 CATATATAGGGCTACTYTTCGTTACGAAGGTTTTAGTGAGATTATGGTAACCCTTTCCAA 2452
          ||| |   | |  || :| || || |||||||||||| | || ||||  || |||   ||
Db    605 CATCTTCCGAGGAACACTACGCTATGAAGGTTTTAGTCAAATAATGGGGACACTTGTGAA 664

Qy   2453 AACTGGGTTCTTTGATGCTGCAAATCATCCACTGCTGCAAGATACTAGTCGTCCAACATA 2512
           |  || |||||   | | | |  |  ||   | ||  | |||   |   | || ||  |
Db    665 GATAGGATTCTTCAGTACAGAATCTACTCTGATTCTTAAGGATGGGATCAGACCCACGCA 724

Qy   2513 TA 2514
           |
Db    725 CA 726


Search completed: June  1, 2002, 00:56:11
Job time: 7632 sec
```

```
                      GenCore version 4.5
              Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          May 31, 2002, 15:45:23 ; Search time 18.01 Seconds
                                             (without alignments)
                                             1386.061 Million cell updates/sec

Title:           US-09-049-304A-122
Perfect score:   5286
Sequence:        1 CARLLLGGGKNGPRVNRIIV..........VPALEILESSGIKLVEKVET 1022

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters:      231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                     %
Result              Query
  No.     Score    Match Length DB  ID                       Description
---------------------------------------------------------------------------
    1       270     5.1      74  1  US-08-474-633A-105        Sequence 105, App
    2     263.5     5.0     123  1  US-08-474-633A-104        Sequence 104, App
    3     154.5     2.9     370  2  US-08-360-606B-32         Sequence 32, Appl
    4       154     2.9     372  2  US-08-360-606B-31         Sequence 31, Appl
    5     128.5     2.4     382  2  US-08-360-606B-30         Sequence 30, Appl
    6     125.5     2.4    1104  4  US-09-268-347-28          Sequence 28, Appl
    7     125.5     2.4    1104  4  US-09-268-347-34          Sequence 34, Appl
    8       112     2.1     476  4  US-09-171-969-4           Sequence 4, Appli
    9     110.5     2.1    1177  3  US-08-855-160-6           Sequence 6, Appli
   10     110.5     2.1    2254  2  US-08-677-010-3           Sequence 3, Appli
   11     110.5     2.1    2254  2  US-08-790-519-3           Sequence 3, Appli
   12     109.5     2.1    1177  3  US-08-855-160-2           Sequence 2, Appli
   13     109.5     2.1    1177  3  US-08-855-160-8           Sequence 8, Appli
   14     109.5     2.1    1288  1  US-07-727-814B-2          Sequence 2, Appli
   15     109.5     2.1    1288  1  US-08-258-614-2           Sequence 2, Appli
   16       109     2.1     363  1  US-08-096-182A-6          Sequence 6, Appli
   17       109     2.1     363  1  US-08-877-109-6           Sequence 6, Appli
   18       109     2.1     363  3  US-08-798-760-6           Sequence 6, Appli
   19       109     2.1     363  5  PCT-US94-08327-6          Sequence 6, Appli
   20       109     2.1     470  3  US-08-879-565-14          Sequence 14, Appl
   21       109     2.1     623  4  US-09-104-068-4           Sequence 4, Appli
   22       109     2.1     637  4  US-09-104-068-2           Sequence 2, Appli
   23     108.5     2.1    1881  4  US-09-233-086-3           Sequence 3, Appli
   24       107     2.0    1022  3  US-08-772-270A-2          Sequence 2, Appli
   25       107     2.0    1176  1  US-08-446-486-6           Sequence 6, Appli
   26       107     2.0    1176  1  US-08-157-363A-10         Sequence 10, Appl
   27       107     2.0    1176  1  US-08-463-308-6           Sequence 6, Appli
```

```
   28       107     2.0    1176  5  PCT-US93-11405A-10        Sequence 10, Appl
   29       107     2.0    1176  5  PCT-US93-11405-10         Sequence 10, Appl
   30       106     2.0     718  2  US-08-974-565C-7          Sequence 7, Appli
   31       106     2.0     718  3  US-09-255-748-7           Sequence 7, Appli
   32       106     2.0    1938  4  US-09-514-302-2           Sequence 2, Appli
   33       105     2.0     833  4  US-09-514-302-3           Sequence 3, Appli
   34     104.5     2.0    1184  6  5254799-6                Patent No. 5254799
   35     102.5     1.9     505  1  US-08-068-395A-3          Sequence 3, Appli
   36     102.5     1.9     505  1  US-08-464-365-3           Sequence 3, Appli
   37     102.5     1.9     893  1  US-07-977-434-4           Sequence 4, Appli
   38     102.5     1.9     893  1  US-08-458-819-4           Sequence 4, Appli
   39     102.5     1.9     893  4  US-09-105-697-10          Sequence 10, Appl
   40     102.5     1.9     893  5  PCT-US91-07035-4          Sequence 4, Appli
   41     102.5     1.9    1140  4  US-09-220-081-2           Sequence 2, Appli
   42     101.5     1.9    1250  3  US-08-938-291A-9          Sequence 9, Appli
   43       101     1.9    2366  1  US-08-480-604A-10         Sequence 10, Appl
   44       101     1.9    2366  2  US-08-405-496A-10         Sequence 10, Appl
   45       101     1.9    2366  4  US-08-915-136-10          Sequence 10, Appl

                              ALIGNMENTS

RESULT   1
US-08-474-633A-105
; Sequence 105, Application US/08474633A
; Patent No. 5773691
;  GENERAL INFORMATION:
;    APPLICANT:  E. I. DU PONT DE NEMOURS AND
;    APPLICANT:  COMPANY
;    TITLE OF INVENTION:  CHIMERIC GENES AND
;    TITLE OF INVENTION:  METHODS FOR INCREASING
;    TITLE OF INVENTION:  INCREASING THE LYSINE
;    TITLE OF INVENTION:  AND THREONINE CONTENT
;    TITLE OF INVENTION:  OF THE SEEDS OF PLANTS
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS
;      ADDRESSEE:  AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  MICROSOFT WORD VERSION 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,633A
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BARBARA C. SIEGELL
;      REGISTRATION NUMBER:  30,684
;      REFERENCE/DOCKET NUMBER:  BB-1037-C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-992-4931
;      TELEFAX:  302-773-0164
;      TELEX:  835420
;  INFORMATION FOR SEQ ID NO:  105:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  74 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-474-633A-105

  Query Match             5.1%;  Score 270;  DB 1;  Length 74;
  Best Local Similarity   67.6%;  Pred. No. 9.9e-19;
```

```
  Matches   50;  Conservative   18;  Mismatches    6;  Indels    0;  Gaps    0;

Qy    948 KHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALE 1007
          || ||||||| ::||::||||| |||||||||||||:::|::|:||:|||: |:|:|||:
Db      1 KHTATLLEFGDIKNGQTTTAMAKTVGIPAAIGALLLIEDKIKTRGVLRPLEAEVYLPALD 60

Qy   1008 ILESSGIKLVEKVE 1021
          ||:: ||||:|| |
Db     61 ILQAYGIKLMEKAE 74


RESULT   2
US-08-474-633A-104
; Sequence 104, Application US/08474633A
; Patent No. 5773691
;  GENERAL INFORMATION:
;    APPLICANT:  E. I. DU PONT DE NEMOURS AND
;    APPLICANT:  COMPANY
;    TITLE OF INVENTION:  CHIMERIC GENES AND
;    TITLE OF INVENTION:  METHODS FOR INCREASING
;    TITLE OF INVENTION:  INCREASING THE LYSINE
;    TITLE OF INVENTION:  AND THREONINE CONTENT
;    TITLE OF INVENTION:  OF THE SEEDS OF PLANTS
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS
;      ADDRESSEE:  AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  MICROSOFT WORD VERSION 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,633A
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BARBARA C. SIEGELL
;      REGISTRATION NUMBER:  30,684
;      REFERENCE/DOCKET NUMBER:  BB-1037-C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-992-4931
;      TELEFAX:  302-773-0164
;      TELEX:  835420
;  INFORMATION FOR SEQ ID NO:  104:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  123 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-474-633A-104

  Query Match             5.0%;  Score 263.5;  DB 1;  Length 123;
  Best Local Similarity   51.6%;  Pred. No. 1.1e-17;
  Matches   64;  Conservative   17;  Mismatches   26;  Indels   17;  Gaps    4;

Qy    526 KIGKVNEYETDVTIDKGGPK------ILILGAGRVCRPAAEFLASYPDICT-------YG 572
          ||||| : | ::   |  |:      :|||||||| ||||:||||   |  :       :|
Db      1 KIGKVQQ-ENEI---KEKPEMTKKSGVLILGAGRVXRPAADFLASVRTISSQQWYKTYFG 56

Qy    573 VDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLPA 632
           |  :   :|||||||| |||:|||:|| :  | :|||:|  ||   ||||:||:|||||
Db     57 ADSEEKTDVHVIVASLYLKDAKETVEGISDVEAVRLDVSDSESLLKYVSQVDVVLSLLPA 116

Qy    633 SFHA 636
          | ||
```

```
Db    117 SCHA 120

RESULT   3
US-08-360-606B-32
; Sequence 32, Application US/08360606B
; Patent No. 5919617
;  GENERAL INFORMATION:
;    APPLICANT:  Jnanendra K. Bhattacharjee
;    APPLICANT:  Richard C. Garrad
;    APPLICANT:  Paul L. Skatrud
;    APPLICANT:  Robert P. Peery
;    TITLE OF INVENTION:  Methods and Reagents for
;    TITLE OF INVENTION:  Detecting Fungal Pathogens in a
;    TITLE OF INVENTION:  Biological Sample
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 S. Wacker Drive Suite 3200
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  U.S.A.
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  MS Word 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/360,606B
;      FILING DATE:  December 21, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berghoff, Paul H.
;      REGISTRATION NUMBER:  30,243
;      REFERENCE/DOCKET NUMBER:  94,319
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312)913-0001
;      TELEFAX:  (312)913-0002
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  370 amino acid residues
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  protein
;    HYPOTHETICAL:  Yes
;    ORIGINAL SOURCE:
;      ORGANISM:  Yarrowia lypolytica
US-08-360-606B-32

  Query Match             2.9%;  Score 154.5;  DB 2;  Length 370;
  Best Local Similarity   19.6%;  Pred. No. 2.9e-06;
  Matches   78;  Conservative   64;  Mismatches  162;  Indels   93;  Gaps   12;

Qy     18 IIVQPSTRRIHHDAQYEDAGCEISED------LSECGLIIGIKQPKLQMILSDRAYAFFS 71
          : |: |  ||  | :: | |  :  :        |  :|||:|:   :       :  |:
Db     37 VFVEKSPLRIFDDQEFVDVGATLVVEEGSWVSAPEDRMIIGLKELPEESFPLSHEHIQFA 96

Qy     72 HTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRY 131
          | :| |     :| :      :|:| | :  |:|:|  ||   || ||
Db     97 HCYKDQGGWKDVLSRFPAGNGTLYDLEFLEDDNGRRVAAFEFHAGFAG------------ 144

Qy    132 LSLGYSTPFLSLGQSH----------MYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFV 181
           ::|  |   :  |:|           ||:       |  : :::|         | | |
Db    145 AAIGVET--WAFQQTHPDSENLPGVSAYPN---ETELVDKIKKDLAAAVEKGSKLPTVLV 199

Qy    182 FTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDI 241
             :|    || :: : :         :||     |: :   :  :
Db    200 IGALGRCGSGAIDLARKV--------GIPE----ENIIRWDMNETK-------------- 233

Qy    242 VSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLM 301
```

```
                       ||  :           : ||  | : :||:|  :  || :| | | :
Db      234 -----------KGGPF-----------DEIAD-ADIFINCIYLSQPIPPFINYDLLNKET 270

Qy      302 ETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDIL 361
                 :| |   | :    :     :|: : |    : :         |  :   ::| |
Db      271 RKLSVIVDVSADTTNPHNPVPVYTIATTFDHPTVPVETT--------AGPKLSVCSIDHL 322

Qy      362 PTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLR 398
            |:   :|||: |      |: ||  : |    | : |
Db      323 PSLLPREASEAFS---EALLPSLLQLPQRDTAPVWTR 356


RESULT   4
US-08-360-606B-31
; Sequence 31, Application US/08360606B
; Patent No. 5919617
;  GENERAL INFORMATION:
;    APPLICANT:  Jnanendra K. Bhattacharjee
;    APPLICANT:  Richard C. Garrad
;    APPLICANT:  Paul L. Skatrud
;    APPLICANT:  Robert P. Peery
;    TITLE OF INVENTION:  Methods and Reagents for
;    TITLE OF INVENTION:  Detecting Fungal Pathogens in a
;    TITLE OF INVENTION:  Biological Sample
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 S. Wacker Drive Suite 3200
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  U.S.A.
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  MS Word 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/360,606B
;      FILING DATE:  December 21, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berghoff, Paul H.
;      REGISTRATION NUMBER:  30,243
;      REFERENCE/DOCKET NUMBER:  94,319
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312)913-0001
;      TELEFAX:  (312)913-0002
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  372  amino acid residues
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  protein
;    HYPOTHETICAL:  Yes
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharomyces cerevisiae
US-08-360-606B-31

  Query Match             2.9%;  Score 154;  DB 2;  Length 372;
  Best Local Similarity   21.4%;  Pred. No. 3.3e-06;
  Matches   85;  Conservative   57;  Mismatches  162;  Indels   94;  Gaps   12;

Qy       17 RIIVQPSTRRIHHDAQYEDAGCEIS-----EDLSECGLIIGIKQ-PKLQMILSDRAYAFF 70
            :| |: | :   :  :|  ||  |      :      :|||:|: |:         :  |
Db       35 KIYVEDSPQSTFNINEYRQAGAIIVPAGSWKTAPRDRIIIGLKEMPETDTFPLVHEHIQF 94

Qy       71 SHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQR 130
            :| :| |     :| : ::   :|:| | :  | |:|  ||| :|| ||       || |
Db       95 AHCYKDQAGWQNVLMRFIKGAGTLYDLEFLENDQGRRVAAFGFYAGFAGA-----ALGVR 149
```

```
Qy      131 YLSLGYSTPFLSLGQSHM----------YPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
                       :  |:|           ||:    || |  | ::           | |
Db      150 ---------DWAFKQTHSDDEDLPAVSPYPN---EKALVKDVTKDYKEALATGARKPTVL 197

Qy      181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRD 240
            :   :|    || ::   :     ||  | :| :                      |||
Db      198 IIGALGRCGSGAIDLLHKV--GIPDANILWDIKE----------------------TSRG 233

Qy      241 IVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQL 300
                      ||:                   | | : :||:|  |   |  ||::|
Db      234 --------GPFDE------------------IPQADIFINCIYLSKPIAPFTNMEKLNNP 267

Qy      301 METGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDI 360
                  :| |   | :    |     :|   :|      :         |  :  :::|
Db      268 NRRLRTVVDVSADTTNPHNPIPIYTVATVFNKPTVLVPTT--------VGPKLSVISIDH 319

Qy      361 LPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLR 398
            ||:   :|||: | :    |: ||  : |    | ::|
Db      320 LPSLLPREASEFFSH---DLLPSLELLPQRKTAPVWVR 354


RESULT   5
US-08-360-606B-30
; Sequence 30, Application US/08360606B
; Patent No. 5919617
;  GENERAL INFORMATION:
;    APPLICANT:  Jnanendra K. Bhattacharjee
;    APPLICANT:  Richard C. Garrad
;    APPLICANT:  Paul L. Skatrud
;    APPLICANT:  Robert P. Peery
;    TITLE OF INVENTION:  Methods and Reagents for
;    TITLE OF INVENTION:  Detecting Fungal Pathogens in a
;    TITLE OF INVENTION:  Biological Sample
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 S. Wacker Drive Suite 3200
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  U.S.A.
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  MS Word 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/360,606B
;      FILING DATE:  December 21, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berghoff, Paul H.
;      REGISTRATION NUMBER:  30,243
;      REFERENCE/DOCKET NUMBER:  94,319
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312)913-0001
;      TELEFAX:  (312)913-0002
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  382  amino acid residues
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  protein
;    HYPOTHETICAL:  Yes
;    ORIGINAL SOURCE:
;      ORGANISM:  Candida albicans
US-08-360-606B-30


  Query Match             2.4%;  Score 128.5;  DB 2;  Length 382;
  Best Local Similarity   20.3%;  Pred. No. 0.001;
  Matches   85;  Conservative   71;  Mismatches  158;  Indels  105;  Gaps   18;
```

```
Qy     18 IIVQPSTRRIHHDAQYEDAGCEISEDLS-----ECGLIIGIKQ-PKLQMILSDRAYAFFS 71
          | |: |::      :||  | :|  : |     :  :| |:|: |: :       :  |:
Db     38 IYVEESSQSTFDIKEYEAVGAKIVPEGSWKTAPKERIIFGLKELPENETFPLIHEHIQFA 97

Qy     72 HTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRA----GLIDF---- 123
          | :| |     :| :  :    |:| | :  | |:|  ||| :|| |    |::|:
Db     98 HCYKDQAGWQDVLKRFPQGNGILYDLEFLENDQGRRVAAFGFYAGFAGAAIGVLDWSFKQ 157

Qy    124 LHG--LGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFV 181
          |:|   | :    |   | ::      ||:       | :  |:  |     |  |   |
Db    158 LNGNTKGTKGEGEGGELPGVT-----PYPNENELIKDVKIELEKALTKN--GGQYPKCLV 210

Qy    182 FTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDI 241
             :|    || ::|| :         :|:     |::|   :
Db    211 IGALGRCGSGAIDLFKKI--------GIPD----DNIAKWDMA----------------- 241

Qy    242 VSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVI-----VNCMYWEKRFPPLLNMDQ 296
                  :  ||                  |:  ::     :||:|  |   | :| :
Db    242 --------ETAKG-----------------GPFQEIVDLDIFINCIYLSKFIFPFINKEI 276

Qy    297 LQQLMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSK--NSYHDDM---EGA 351
          |         |  : |:: |          :|:   |   |: :   | :  ::   :|
Db    277 LN---NENRKLTTIVDVSAD----------TTNPHDPIPVYEIATVFNEFTVEVKLDKGF 323

Qy    352 GVVCLAVDILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACI--AHAGRL 408
           :   ::| ||:   :|||: |      |: ||  :      | ::|   :   |  ||
Db    324 KLSVCSIDHLPSLLPREASEFFA---KDLMPSLLELFNRDTSPVWVRAKQLFDKHVARL 379

RESULT   6
US-09-268-347-28
; Sequence 28, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
;   LENGTH: 1104
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-268-347-28


  Query Match             2.4%;  Score 125.5;  DB 4;  Length 1104;
  Best Local Similarity   22.5%;  Pred. No. 0.013;
  Matches  111;  Conservative   74;  Mismatches  193;  Indels  115;  Gaps   27;

Qy    454 INEALDI-IETAGGSFHLVRCEVGQSTDDMSYSELEVGADDTATL-------------DK 499
          :|  |:   ::  |    :: |    |       ::|| |  ||:              :
Db    521 LNNGLNFKFKSTDGELLNIKVENDTVTFTPKKGSVQVGEDGKATIQNGTKTTDGLVEASE 580

Qy    500 IIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYETDVTIDKGGPKILILGAG---RVCR 556
          :::||  |  : | | |   |::     |  ||     |: | | |:  | ||   :| :
Db    581 LVESLNKLGWKVGVDKDGSGELD-----GASNE-----TLVKSGDKV-TLKAGENLKVKQ 629

Qy    557 PAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEE---------TVDGIENTTATQ 607
              |        || : |         | |:  ||            | ||:  |:|
Db    630 DGTNF--------TYALKDELTG-----VKSVEFKDTANGSNGASTKITKDGLTITSANG 676

Qy    608 LDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASY--VDESMSNLSQA 665
           :    |: :    :::|    : |  : |:  |   |||    |::  :  |  ||::
Db    677 AN----GAAATDADKIKVASDGISAG-NKAVKNVVSGLKK-FGDANFNPLTSSADNLTKQ 730

Qy    666 AKDA--GVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLA 723
            ||  |:| | | | |       :: : |   |  | ::       ||    :|:
Db    731 YDDAYKGLTNLDEKGADK-----QTLTVADNTAATVGDLR-------GLGWVISAD---- 774

Qy    724 YKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSL 783
                  | |    |  | :   |     |: :: | | :  |    | |    : |::
Db    775 -----KTTGELNKEYNAQV-RNANEVKFKSGNGIHVSGKTVNGRREITFEL--AKDENAI 826

Qy    784 IYGDLYGISK--EASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGFL 841
           :|  || ||    :|:   |         :|   |:| |   |:  ::| :  :|
Db    827 AFG--YG-SKALRDNTVAIGTGN-------VVNAEKSGAFGDPNY--IEDKAGGSYAFGN 874

Qy    842 DELLNNISTINT-------DLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLG 894
          |   | |:: ||       :   :|:|  | : :         |:|:    ||   | ||
Db    875 D---NRITSKNTFVLGNGVNAKYKANGDVDTETVT-------VKDKDGKETTVTVPKALG 924

Qy    895 LHEETQIPKGCSS 907
             |  :  |  |
Db    925 ATVENSVYLGNKS 937

RESULT   7
US-09-268-347-34
; Sequence 34, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
;   LENGTH: 1104
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-268-347-34


  Query Match             2.4%;  Score 125.5;  DB 4;  Length 1104;
  Best Local Similarity   22.5%;  Pred. No. 0.013;
  Matches  111;  Conservative   74;  Mismatches  193;  Indels  115;  Gaps   27;

Qy    454 INEALDI-IETAGGSFHLVRCEVGQSTDDMSYSELEVGADDTATL-------------DK 499
          :|  |:   ::  |    :: |    |       ::|| |  ||:              :
Db    521 LNNGLNFKFKSTDGELLNIKVENDTVTFTPKKGSVQVGEDGKATIQNGTKTTDGLVEASE 580

Qy    500 IIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYETDVTIDKGGPKILILGAG---RVCR 556
          :::||  |  : | | |   |::     |  ||     |: | | |:  | ||   :| :
Db    581 LVESLNKLGWKVGVDKDGSGELD-----GASNE-----TLVKSGDKV-TLKAGENLKVKQ 629

Qy    557 PAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEE---------TVDGIENTTATQ 607
              |        || : |         | |:  ||            | ||:  |:|
Db    630 DGTNF--------TYALKDELTG-----VKSVEFKDTANGSNGASTKITKDGLTITSANG 676

Qy    608 LDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASY--VDESMSNLSQA 665
           :    |: :    :::|    : |  : |:  |   |||    |::  :  |  ||::
Db    677 AN----GAAATDADKIKVASDGISAG-NKAVKNVVSGLKK-FGDANFNPLTSSADNLTKQ 730

Qy    666 AKDA--GVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLA 723
            ||  |:| | | | |       :: : |   |  | ::       ||    :|:
Db    731 YDDAYKGLTNLDEKGADK-----QTLTVADNTAATVGDLR-------GLGWVISAD---- 774

Qy    724 YKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSL 783
                  | |    |  | :   |     |: :: | | :  |    | |    : |::
Db    775 -----KTTGELNKEYNAQV-RNANEVKFKSGNGIHVSGKTVNGRREITFEL--AKDENAI 826

Qy    784 IYGDLYGISK--EASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGFL 841
           :|  || ||    :|:   |         :|   |:| |   |:  ::| :  :|
Db    827 AFG--YG-SKALRDNTVAIGTGN-------VVNAEKSGAFGDPNY--IEDKAGGSYAFGN 874

Qy    842 DELLNNISTINT-------DLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLG 894
          |   | |:: ||       :   :|:|  | : :         |:|:    ||   | ||
Db    875 D---NRITSKNTFVLGNGVNAKYKANGDVDTETVT-------VKDKDGKETTVTVPKALG 924

Qy    895 LHEETQIPKGCSS 907
```

```
                 |  :  |  |
Db    925 ATVENSVYLGNKS 937

RESULT   8
US-09-171-969-4
; Sequence 4, Application US/09171969
; Patent No. 6284533
;  GENERAL INFORMATION:
;    APPLICANT:  Thomas, Lawrence J.
;    TITLE OF INVENTION:  PLASMID-BASED VACCINE FOR TREATING ATHEROSCLEROSIS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Witcoff, Ltd.
;      STREET:  75 State Street, Suite 2300
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1807
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/171,969
;      FILING DATE:  01 May 1997 (01.05.97)
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/640,713
;      FILING DATE:  01 May 1996 (01.05.96)
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/802,967
;      FILING DATE:  21 February 1997 (21.02.97)
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Leon R. Yankwich
;      REGISTRATION NUMBER:  30,237
;      REFERENCE/DOCKET NUMBER:  TCS 414.1 PCT (05872)
;   INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  476 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    FEATURE:
;      NAME/KEY:  Amino acid sequence of mature human
;      NAME/KEY:  CETP
;      LOCATION:
;    PUBLICATION INFORMATION:
;      AUTHORS:  Drayna, Dennis, et al.
;      TITLE:  Cloning and sequencing of human
;      TITLE:  cholesteryl ester transfer cDNA
;      JOURNAL:  Nature
;      VOLUME:  327
;      ISSUE:
;      PAGES:  632 - 634
;      DATE:  18-JUN-1987
;      RELEVANT RESIDUES IN SEQ ID NO:  4: FROM 1 TO 476
US-09-171-969-4


  Query Match              2.1%;  Score 112;  DB 4;  Length 476;
  Best Local Similarity   19.0%;  Pred. No. 0.061;
  Matches  111;  Conservative   74;  Mismatches  192;  Indels  208;  Gaps   29;

Qy    512 GGDHDAGQEIELALKIGKVNEYETDVTIDKGGPKILILG--AGRVCRPAAEFLASYPDIC 569
          |  |:||    :  :| |              | :|:|     :| : | :  ||||||
Db      4 GTSHEAG----IVCRITK--------------PALLVLNHETAKVIQTAFQ-RASYPDI- 43

Qy    570 TYGVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISL 629
                   ::  :::  :          |: |   : | :|         ||||
Db     44 -------TGEKAMMLLGQVKY--------GLHNIQISHLSIAS--------SQVE----- 75

Qy    630 LPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDA---GVTILCEMGLDPG---- 682
                             :| |  :| |: |:|   |     | |    :|:|
Db     76 -------------------LVEAKSIDVSIQNVSVVFKGTLKYGYTTAWWLGIDQSIDFE 116

Qy    683 IDHLMSMKMIDEAHARKGKIKAFTSYC-----------GGLPSPA--------------- 716
          ||  : :::  :     |:::     |            |   |
Db    117 IDSAIDLQINTQLTCDSGRVRTDAPDCYLSFHKLLLHLQGEREPGWIKQLFTNFISFTLK 176

Qy    717 ------------AANNPLAYKFSWNPAGALRSGKNPAVYKFLGETI--------HVDGHN 756
                        :| :|       |  |  |         |: :        |  ||
Db    177 LVLKGQICKEINVISNIMADFVQTRAASILSDGDIGVDISLTGDPVITASYLESHHKGHF 236

Qy    757 LYESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSE-IMVTL 815
          :|::       :||      ||  :  : ||             :   |  ||| :  :|
Db    237 IYKNVS----EDLP------LPTFSPTLLGD-------------SRMLYFWFSERVFHSL 273

Qy    816 SKTGFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDI--EASGGYDD--DLIARL 871
          :|  | |     |:       :|  |:         ||: :|  |  ||:     :
Db    274 AKVAFQDGR---LMLSLMGDEFKAVLETW-----GFNTNQEIFQEVVGGFPSQAQVTVHC 325

Qy    872 LKLG--CCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHN-EQD 928
          ||:    |:|| : | :   :|||                     | :|: :  :  |:|
Db    326 LKMPKISCQNKGVVVNSSVMVKFL-------------------FPRPDQQHSVAYTFEED 366

Qy    929 MVLLHHEVEVEYP---------DGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIG 979
          :|     |:  |          | |   |  : | |           :|   ||||  :
Db    367 IVT---TVQASYSKKKLFLSLLDFQITPKTVSNLTESSSESIQSFLQSMITAVGIPEVMS 423

Qy    980 AL-LLLKNKVQTKGV-----IRPLQPEIYVPALEILESSGIKLVE 1018
           | ::    : :|||     | |          ||:   |  |::
Db    424 RLEVVFTALMNSKGVSLFDIINP----------EIITRDGFLLLQ 458


RESULT   9
US-08-855-160-6
; Sequence 6, Application US/08855160
; Patent No. 6090931
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, David L.
;    APPLICANT:  Herrnstadt, Corinna
;    APPLICANT:  Wilcox, Edward R.
;    APPLICANT:  Wong, Siu-Yin
;    TITLE OF INVENTION:  Process for Altering the Host Range of
;    TITLE OF INVENTION:  Bacillus Thuringiensis Toxins, and No. 6090931el Toxins Pro
;    TITLE OF INVENTION:  Thereby
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik A
;      ADDRESSEE:  Professional Association
;      STREET:  2421 NW 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606-6669
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/855,160
;      FILING DATE:  13-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/580,781
;      FILING DATE:  29-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/420,615
;      FILING DATE:  10-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/097,808
```

```
;      FILING DATE:  27-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/980,128
;      FILING DATE:  23-NOV-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/808,129
;      FILING DATE:  13-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/356,599
;      FILING DATE:  24-MAY-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/904,572
;      FILING DATE:  05-SEP-1986
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/808,129
;      FILING DATE:  12-DEC-1985
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lloyd, Jeff
;      REGISTRATION NUMBER:  35,589
;      REFERENCE/DOCKET NUMBER:  M12C1FDF3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (352) 375-8100
;      TELEFAX:  (352) 372-5800
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1177 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-855-160-6

  Query Match             2.1%;  Score 110.5;  DB 3;  Length 1177;
  Best Local Similarity   21.3%;  Pred. No. 0.43;
  Matches  180;  Conservative  100;  Mismatches  311;  Indels  253;  Gaps   44;

Qy     336 RYDPSKNSYHDDMEGAGVVCLAVDIL--PTEFSKEASQHFGNIL-------------SRL 380
           | :  |    |  |  :|    :| |:  ||   :  |  | ||:             : |
Db     423 RQEFSHRLSHVSMFRSGFSNSSVSIIRAPTFSWQHRSAEFNNIIPSSQITQIPLTKSTNL 482

Qy     381 VASLASVKQPAEL-PSYLRRACIAHAGRL-----TPLYE-YIPRMR---------NTMID 424
            :  : || |       |||        |      || : |  |:|         :| ||
Db     483 GSGTSVVKGPGFTGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSID 542

Qy     425 LAPAK------------------------TNPLPDKKYSTLVSLSGHLFDKFLINEA-LD 459
             |                          | |      |:: :|| |:|:    ||  :|
Db     543 GRPINQGNFSATMSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSG--NEVYID 600

Qy     460 IIE------TAGGSFHLVRCEVGQSTDDMSYSELEVGADDTAT---LDKIIDSLTSLANE 510
            ||      |    : | | :  :: :::  |  ::|     |   :|:: : :  |::|
Db     601 RIEFVPAEVTFEAEYDLERAQ--KAVNELFTSSNQIGLKTDVTDYHIDQVSNLVECLSDE 658

Qy     511 HGGDHDAGQEIELALKIGK----------------VNEY-------ETDVTIDKGGPKIL 547
              |    ||:   :|  |                :|          ||:|| :||  :
Db     659 FCLDEK--QELSEKVKHAKRLSDERNLLQDPNFRGINRQLDRGWRGSTDITI-QGGDDVF 715

Qy     548 ------ILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDG-- 599
                 :||    | |            ||                ||||  |  :
Db     716 KENYVTLLGTFDECYP------------TY----------------LYQKIDESKLKAYT 747

Qy     600 -------IENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKK---HM 649
                  ||::   :: :    :  : |: |    || | |  : | | | |  :   |:
Db     748 RYQLRGYIEDSQDLEIYLIRYNAKHETVN-VPGTGSLWPLSAQSPI-GKCGEPNRCAPHL 805

Qy     650 VTASYVDESMSNLSQAAK---------DAGVTILCEMGLDPGIDHLMSMKMIDEAHARKG 700
                :| |  :  : |          | | | | |   | |:  :  :|  |  ||| |
Db     806 EWNPDLDCSCRDGEKCAHHSHHFSLDIDVGCTDLNE---DLGVWVIFKIKTQD-GHARLG 861

Qy     701 KIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHN---- 756
            ::        |   | |    | |  |         :   |||   |::     |
Db     862 NLEFLEE--KPLVGEALARVKRAEK-KWRDKREKLEWETNIVYKEAKESVDALFVNSQYD 918
```

```
Qy     757 ----------LYESAKRLR-LRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRY 805
                     :: :  ||:  :||     |  :|  |: |: :| |       |: |   |
Db     919 QLQADTNIAMIHAADKRVHSIREAYLPELSVIPGVNAAIFEELEG------RIFTAFSLY 972

Qy     806 EGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGFLD-ELLNN----------ISTINTD 854
           :  : |     | | |:          |    || :| |  ||           : :: :
Db     973 DARNVI-----KNGDFN-------NGLSCWNVKGHVDVEEQNNQRSVLVLPEWEAEVSQE 1020

Qy     855 LDIEASGGYDDDLIARLLKL------GCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSP 908
           : :    ||    || :         ||    ||   |   :||    ||   |    :
Db    1021 VRVCPGRGY----ILRVTAYKEGYGEGCVTIHEIENNT-DELKFSNCVEEEIYPNNTVTC 1075

Qy     909 FDVICQRMEQRMAY-----GHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEFGKVENGR 963
            |    : |   ||     |:||   |              ||:   |:: |     :||
Db    1076 NDYTVNQEEYGGAYTSRNRGYNEAPSV--------------PAD--YASVYEEKSYTDGR 1119

Qy     964 STTAMALTVG----IPAAIGALLLLKNKVQTKGVIRPLQ--PEIYVPALEILESSGIKLV 1017
                    |     |  :|             | : |:  ||     :|| |: |  :|
Db    1120 RENPCEFNRGYRDYTPLPVGY------------VTKELEYFPETDKVWIEIGETEGTFIV 1167

Qy    1018 EKVE 1021
           : ||
Db    1168 DSVE 1171


RESULT  10
US-08-677-010-3
; Sequence 3, Application US/08677010
; Patent No. 5925805
;  GENERAL INFORMATION:
;    APPLICANT:  Ohlrogge, John B.
;    APPLICANT:  Roesler, Keith R.
;    APPLICANT:  Shorrosh, Basil S.
;    TITLE OF INVENTION:  Structure and Expression of an
;    TITLE OF INVENTION:  Arabidopsis Acetyl-coenzyme A Carboxylase Gene
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/677,010
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  6550-00002CPA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810)641-1600
;      TELEFAX:  (810)641-0270
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2254 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-677-010-3

  Query Match             2.1%;  Score 110.5;  DB 2;  Length 2254;
  Best Local Similarity   18.6%;  Pred. No. 1.4;
  Matches  170;  Conservative  110;  Mismatches  308;  Indels  325;  Gaps   42;
```

```
Qy     131 YLSLGYSTP-FLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVS 189
           ||  |   |  :||  | :  ::  :|  : ||    .|: |               | |
Db     579 YLEKGQIPPKHISLVHSQVSLNIEGSKYTIDVVRGGSGTYRLRM-------------NKS 625

Qy     190 QGAQEIFKLLPHTFVDAEKLPEIFQARNL---SKQSQSTKRVFQLYGCVVTSRDIVSHKD 246
           :   ||     ||  |   | ::    ::    :::  |: :     |:     : :  |
Db     626 EVVAEI-----HTLRDGGLLMQLDGKSHVIYAEEEAAGTRLLIDGRTCL-----LQNDHD 675

Qy     247 PTRQFDKGDYYAHPEHYTPVFHERI-----------APYASVIV--NCMYWEKRFPPLLN 293
           |::   :          ||    |             ||| | |   ||       |||:
Db     676 PSKLMAE----------TPCKLMRYLISDNSNIDADTPYAEVEVMKMCM-------PLLS 718

Qy     294 MDQLQQLMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPS----KNSYHDDME 349
                       |  ||       | :::      :::      |||       :|
Db     719 ------------PASGVIHFKMSEGQAMQAGELIANLDLD----DPSAVRKAEPFHGSFP 762

Qy     350 GAGVVCLAVDILPTEFSKEASQH--------------FGNILSRLVASLASVKQPAELPS 395
             |        |||  |    |               : : :  :|  | :     |||
Db     763 RLG--------LPTAISGRVHQRCAATLNAARMILAGYEHKVDEVVQDLLNCLDSPELPF 814

Qy     396 YLRRACIAHAGRLTPLYEYIPRMRNTMIDLAPAKTNPLPDKKYSTLVSLS-GHLFDKFLI 454
              : | |      |       :|| :            : ||    |:|   |   |
Db     815 LQWQECFAVLATRLP-----KNLRNML------------ESKYREFESISRNSLTTDFPA 857

Qy     455 NEALDIIETAGGSFHLVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGD 514
                |:|      ||  |:           | | ||     |:::|: | |||  : |
Db     858 KLLKGILEA-----HLSSCD-----------EKERGA-----LERLIEPLMSLAKSYEG- 895

Qy     515 HDAGQEIELALKIGKVNEYETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICTYGVD 574
              |:|    : :  : |                           |:| |  ::     |
Db     896 ---GRESHARVIVHSLFE---------------------------EYL-SVEELFN---D 921

Qy     575 DHDADQIHVIVASLYQKDAEETVD------GIEN-------------------------- 602
           :  || |   :  ||:||  : ||      ||:|
Db     922 NMLADVIE-RMRQLYKKDLLKIVDIVLSHQGIKNKNKLVLRLMEQLVYPNPAAYRDKLIR 980

Qy     603 -TTATQLDVADIG-SLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVD---- 656
            :|    : :::    | |: | ::      :   : ||    ||:        :|
Db     981 FSTLNHTNYSELALKASQLLEQTKL------SELRSNIARSLSELEMFTEDGENMDTPKR 1034

Qy     657 --------ESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSY 708
                   | : : | | :|| |      ||    || :  ::::             :|
Db    1035 KSAINERIEDLVSASLAVEDALV------GLFDHSDHTLQRRVVE-------------TY 1075

Qy     709 CGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRE 768
              |  |    :  : :  |: :| | | :
Db    1076 IRRLYQPYVVKD--SVRMQWHRSGLLASWE------------------------------ 1103

Qy     769 LPAFALEHLPNRNSLIYGDLYGI----SKEASTIYRATXRYEGFSEIMVTLSKTGFFDAA 824
              |  ||:  :|        |:    : |   : : : |  |    || :    |  :
Db    1104 ---FLEEHMERKN-------IGLDDPDTSEKGLVEKRSKRKWG---AMVIIKSLQFLPSI 1150

Qy     825 NHPLLQDTSRPTYKGFLDELLNN-----ISTINTDLDIEASGGYDDDLIARLLKLGCCKN 879
               |::|    |:     |  |     |  ||  : :    | :|    |: ||
Db    1151 ISAALRETKHNDYETAGAPLSGNMMHIAIVGINNQMSLLQDSGDEDQAQERVNKL----- 1205

Qy     880 KEIAVKTVKTIKFLGLHEETQIPKGCSSPFDV---ICQRMEQRMAYGHNEQDMVLLHHEV 936
                   | :|     ||      ||:   |   | || | |    |:       |  :
Db    1206 -------AKILK-----EEEVSSSLCSAGVGVISCIIQRDEGRTPMRHS------FHWSL 1247

Qy     937 EVEYPDGQPAEKH 949
           | :|   :|  :|
Db    1248 EKQYYVEEPLLRH 1260

RESULT 11
US-08-790-519-3
; Sequence 3, Application US/08790519
; Patent No. 5962767
;  GENERAL INFORMATION:
;    APPLICANT:  Ohlrogge, John B.
;    APPLICANT:  Roesler, Keith R.
;    APPLICANT:  Shorrosh, Basil S.
;    TITLE OF INVENTION:  Structure and Expression of an
;    TITLE OF INVENTION:  Arabidopsis Acetyl-coenzyme A Carboxylase Gene
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  Michigan
;      COUNTRY:  U.S.A.
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/790,519
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/248,630
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  6550-00002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810)641-1600
;      TELEFAX:  (810)641-0270
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2254 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-790-519-3

  Query Match             2.1%;  Score 110.5;  DB 2;  Length 2254;
  Best Local Similarity   18.6%;  Pred. No. 1.4;
  Matches  170;  Conservative  110;  Mismatches  308;  Indels  325;  Gaps   42;

Qy     131 YLSLGYSTP-FLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVS 189
           ||  |   |  :||  | :  ::  :|  : ||    .|: |               | |
Db     579 YLEKGQIPPKHISLVHSQVSLNIEGSKYTIDVVRGGSGTYRLRM-------------NKS 625

Qy     190 QGAQEIFKLLPHTFVDAEKLPEIFQARNL---SKQSQSTKRVFQLYGCVVTSRDIVSHKD 246
           :   ||     ||  |   | ::    ::    :::  |: :     |:     : :  |
Db     626 EVVAEI-----HTLRDGGLLMQLDGKSHVIYAEEEAAGTRLLIDGRTCL-----LQNDHD 675

Qy     247 PTRQFDKGDYYAHPEHYTPVFHERI-----------APYASVIV--NCMYWEKRFPPLLN 293
           |::   :          ||    |             ||| | |   ||       |||:
Db     676 PSKLMAE----------TPCKLMRYLISDNSNIDADTPYAEVEVMKMCM-------PLLS 718

Qy     294 MDQLQQLMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPS----KNSYHDDME 349
                       |  ||       | :::      :::      |||       :|
Db     719 ------------PASGVIHFKMSEGQAMQAGELIANLDLD----DPSAVRKAEPFHGSFP 762

Qy     350 GAGVVCLAVDILPTEFSKEASQH--------------FGNILSRLVASLASVKQPAELPS 395
             |        |||  |    |               : : :  :|  | :     |||
Db     763 RLG--------LPTAISGRVHQRCAATLNAARMILAGYEHKVDEVVQDLLNCLDSPELPF 814

Qy     396 YLRRACIAHAGRLTPLYEYIPRMRNTMIDLAPAKTNPLPDKKYSTLVSLS-GHLFDKFLI 454
              : | |      |       :|| :            : ||    |:|   |   |
Db     815 LQWQECFAVLATRLP-----KNLRNML------------ESKYREFESISRNSLTTDFPA 857

Qy     455 NEALDIIETAGGSFHLVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGD 514
                |:|      ||  |:           | | ||     |:::|: | |||  : |
Db     858 KLLKGILEA-----HLSSCD-----------EKERGA-----LERLIEPLMSLAKSYEG- 895

Qy     515 HDAGQEIELALKIGKVNEYETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICTYGVD 574
              |:|    : :  : |                           |:| |  ::     |
```

```
Db    896 ---GRESHARVIVHSLFE--------------------------EYL-SVEELFN---D 921

Qy    575 DHDADQIHVIVASLYQKDAEETVD------GIEN-------------------------- 602
          :  || |   :  ||:||  : ||      ||:|
Db    922 NMLADVIE-RMRQLYKKDLLKIVDIVLSHQGIKNKNKLVLRLMEQLVYPNPAAYRDKLIR 980

Qy    603 -TTATQLDVADIG-SLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVD---- 656
           :|    : :::    | |: | ::      :   : ||    ||:        :|
Db    981 FSTLNHTNYSELALKASQLLEQTKL------SELRSNIARSLSELEMFTEDGENMDTPKR 1034

Qy    657 --------ESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSY 708
                  | : : | | :|| |      ||    || :  ::::             :|
Db   1035 KSAINERIEDLVSASLAVEDALV------GLFDHSDHTLQRRVVE------------TY 1075

Qy    709 CGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRE 768
             |  |    :  : :  |: :| | | :
Db   1076 IRRLYQPYVVKD--SVRMQWHRSGLLASWE----------------------------- 1103

Qy    769 LPAFALEHLPNRNSLIYGDLYGI----SKEASTIYRATXRYEGFSEIMVTLSKTGFFDAA 824
             |  ||:  :|        |:    : |   : : :  |  |    || :    |  :
Db   1104 ---FLEEHMERKN-------IGLDDPDTSEKGLVEKRSKRKWG---AMVIIKSLQFLPSI 1150

Qy    825 NHPLLQDTSRPTYKGFLDELLNN-----ISTINTDLDIEASGGYDDDLIARLLKLGCCKN 879
              |::|    |:      |  |     |  ||  : :    | :|    |: ||
Db   1151 ISAALRETKHNDYETAGAPLSGNMMHIAIVGINNQMSLLQDSGDEDQAQERVNKL----- 1205

Qy    880 KEIAVKTVKTIKFLGLHEETQIPKGCSSPFDV---ICQRMEQRMAYGHNEQDMVLLHHEV 936
                  | :|     ||      ||:   |   | || | |    |:       |  :
Db   1206 -------AKILK-----EEEVSSSLCSAGVGVISCIIQRDEGRTPMRHS------FHWSL 1247

Qy    937 EVEYPDGQPAEKH 949
          | :|   :|  :|
Db   1248 EKQYYVEEPLLRH 1260

RESULT  12
US-08-855-160-2
; Sequence 2, Application US/08855160
; Patent No. 6090931
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, David L.
;    APPLICANT:  Herrnstadt, Corinna
;    APPLICANT:  Wilcox, Edward R.
;    APPLICANT:  Wong, Siu-Yin
;    TITLE OF INVENTION:  Process for Altering the Host Range of
;    TITLE OF INVENTION:  Bacillus Thuringiensis Toxins, and No. 6090931el Toxins Produc
;    TITLE OF INVENTION:  Thereby
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik A
;      ADDRESSEE:  Professional Association
;      STREET:  2421 NW 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606-6669
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/855,160
;      FILING DATE:  13-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/580,781
;      FILING DATE:  29-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/420,615
;      FILING DATE:  10-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/097,808
;      FILING DATE:  27-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/980,128
;      FILING DATE:  23-NOV-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/808,129
;      FILING DATE:  13-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/356,599
;      FILING DATE:  24-MAY-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/904,572
;      FILING DATE:  05-SEP-1986
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/808,129
;      FILING DATE:  12-DEC-1985
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lloyd, Jeff
;      REGISTRATION NUMBER:  35,589
;      REFERENCE/DOCKET NUMBER:  M12C1FDF3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (352) 375-8100
;      TELEFAX:  (352) 372-5800
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1177 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-855-160-2

  Query Match             2.1%;  Score 109.5;  DB 3;  Length 1177;
  Best Local Similarity   21.3%;  Pred. No. 0.54;
  Matches  178;  Conservative   99;  Mismatches  305;  Indels  253;  Gaps   44;

Qy    345 HDDMEGAGVVCLAVDIL--PTEFSKEASQHFGNIL-------------SRLVASLASVKQ 389
          |  |  :|    :| |:  ||   :  |  | ||:             : | :  : ||
Db    432 HVSMFRSGFSNSSVSIIRAPTFSWQHRSAEFNNIIPSSQITQIPLTKSTNLGSGTSVVKG 491

Qy    390 PAEL-PSYLRRACIAHAGRL-----TPLYE-YIPRMR---------NTMIDLAPAK---- 429
          |       |||        |      || : |  |:|         :| ||  |
Db    492 PGFTGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNF 551

Qy    430 -------------------TNPLPDKKYSTLVSLSGHLFDKFLINEA-LDIIE------ 462
                              | |      |:: :|| |:|:    ||  :| ||
Db    552 SATMSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSG--NEVYIDRIEFVPAEV 609

Qy    463 TAGGSFHLVRCEVGQSTDDMSYSELEVGADDTAT---LDKIIDSLTSLANEHGGDHDAGQ 519
          |    : | | :  :: :::  |  ::|     |   :|:: : :  |::|   |    |
Db    610 TFEAEYDLERAQ--KAVNELFTSSNQIGLKTDVTDYHIDQVSNLVECLSDEFCLDEK--Q 665

Qy    520 EIELALKIGK----------------VNEY-------ETDVTIDKGGPKIL------ILG 550
          |:   :|  |                :|          ||:|| :||  :       :||
Db    666 ELSEKVKHAKRLSDERNLLQDPNFRGINRQLDRGWRGSTDITI-QGGDDVFKENYVTLLG 724

Qy    551 AGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDG---------IE 601
              | |            ||                ||||  |  :           ||
Db    725 TFDECYP------------TY----------------LYQKIDESKLKAYTRYQLRGYIE 756

Qy    602 NTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKK---HMVTASYVDES 658
          ::   :: :    :  : |: |    || | |  : | | | |  :   |:     :| |
Db    757 DSQDLEIYLIRYNAKHETVN-VPGTGSLWPLSAQSPI-GKCGEPNRCAPHLEWNPDLDCS 814

Qy    659 MSNLSQAAK---------DAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYC 709
            :  : |          | | | | |   | |:  :  :|  |  ||| | ::
Db    815 CRDGEKCAHHSHHFSLDIDVGCTDLNE---DLGVWVIFKIKTQD-GHARLGNLEFLEE-- 868

Qy    710 GGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHN------------- 756
            |   | |    | |  |         :   |||   |::     |
Db    869 KPLVGEALARVKRAEK-KWRDKREKLEWETNIVYKEAKESVDALFVNSQYDQLQADTNIA 927
```

```
Qy     757 -LYESAKRLR-LRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVT 814
            :: : ||:  :||     |  :|  |: |: :| |       |: |   |:  : |
Db     928 MIHAADKRVHSIREAYLPELSVIPGVNAAIFEELEG------RIFTAFSLYDARNVI--- 978

Qy     815 LSKTGFFDAANHPLLQDTSRPTYKGFLD-ELLNN----------ISTINTDLDIEASGGY 863
             | | |:          |    || :| |  ||           : :: :: :    ||
Db     979 --KNGDFN-------NGLSCWNVKGHVDVEEQNNQRSVLVLPEWEAEVSQEVRVCPGRGY 1029

Qy     864 DDDLIARLLKL------GCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRME 917
               | |:         ||    ||   |   :||    ||   |    :  |    : |
Db    1030 ----ILRVTAYKEGYGEGCVTIHEIENNT-DELKFSNCVEEEIYPNNTVTCNDYTVNQEE 1084

Qy     918 QRMAY-----GHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEFGKVENGRSTTAMALTV 972
              ||     |:||   |              ||:   |:: |     :||
Db    1085 YGGAYTSRNRGYNEAPSV--------------PAD--YASVYEEKSYTDGRRENPCEFNR 1128

Qy     973 G----IPAAIGALLLLKNKVQTKGVIRPLQ--PEIYVPALEILESSGIKLVEKVE 1021
           |     |  :|             | : |:  ||     :|| |: |  :|: ||
Db    1129 GYRDYTPLPVGY------------VTKELEYFPETDKVWIEIGETEGTFIVDSVE 1171


RESULT  13
US-08-855-160-8
; Sequence 8, Application US/08855160
; Patent No. 6090931
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, David L.
;    APPLICANT:  Herrnstadt, Corinna
;    APPLICANT:  Wilcox, Edward R.
;    APPLICANT:  Wong, Siu-Yin
;    TITLE OF INVENTION:  Process for Altering the Host Range of
;    TITLE OF INVENTION:  Bacillus Thuringiensis Toxins, and No. 6090931el Toxins Produc
;    TITLE OF INVENTION:  Thereby
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik A
;      ADDRESSEE:  Professional Association
;      STREET:  2421 NW 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606-6669
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/855,160
;      FILING DATE:  13-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/580,781
;      FILING DATE:  29-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/420,615
;      FILING DATE:  10-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/097,808
;      FILING DATE:  27-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/980,128
;      FILING DATE:  23-NOV-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/808,129
;      FILING DATE:  13-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/356,599
;      FILING DATE:  24-MAY-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 06/904,572
;      FILING DATE:  05-SEP-1986
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 06/808,129
;      FILING DATE:  12-DEC-1985
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lloyd, Jeff
;      REGISTRATION NUMBER:  35,589
;      REFERENCE/DOCKET NUMBER:  M12C1FDF3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (352) 375-8100
;      TELEFAX:  (352) 372-5800
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1177 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-855-160-8

  Query Match             2.1%;  Score 109.5;  DB 3;  Length 1177;
  Best Local Similarity   21.3%;  Pred. No. 0.54;
  Matches  178;  Conservative   99;  Mismatches  305;  Indels  253;  Gaps   44;

Qy     345 HDDMEGAGVVCLAVDIL--PTEFSKEASQHFGNIL-------------SRLVASLASVKQ 389
           |  |  :|    :| |:  ||   :  |  | ||:             : | :  : ||
Db     432 HVSMFRSGFSNSSVSIIRAPTFSWQHRSAEFNNIIPSSQITQIPLTKSTNLGSGTSVVKG 491

Qy     390 PAEL-PSYLRRACIAHAGRL-----TPLYE-YIPRMR---------NTMIDLAPAK---- 429
           |       |||        |      || : |  |:|         :| ||  |
Db     492 PGFTGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNF 551

Qy     430 --------------------TNPLPDKKYSTLVSLSGHLFDKFLINEA-LDIIE------ 462
                               | |      |:: :|| |:|:    ||  :| ||
Db     552 SATMSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSG--NEVYIDRIEFVPAEV 609

Qy     463 TAGGSFHLVRCEVGQSTDDMSYSELEVGADDTAT---LDKIIDSLTSLANEHGGDHDAGQ 519
           |    : | |  :: :::  |  ::|     |   :|:: : :  |::|   |    |
Db     610 TFEAEYDLERAQ--KAVNELFTSSNQIGLKTDVTDYHIDQVSNLVECLSDEFCLDEK--Q 665

Qy     520 EIELALKIGK----------------VNEY-------ETDVTIDKGGPKIL------ILG 550
           |:   :|  |                :|          ||:|| :||  :       :||
Db     666 ELSEKVKHAKRLSDERNLLQDPNFRGINRQLDRGWRGSTDITI-QGGDDVFKENYVTLLG 724

Qy     551 AGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDG---------IE 601
               | |            ||                ||||  |  :           ||
Db     725 TFDECYP------------TY----------------LYQKIDESKLKAYTRYQLRGYIE 756

Qy     602 NTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKK---HMVTASYVDES 658
           ::   :: :    :  : |: |    || | |  : | | | |  :   |:     :| |
Db     757 DSQDLEIYLIRYNAKHETVN-VPGTGSLWPLSAQSPI-GKCGEPNRCAPHLEWNPDLDCS 814

Qy     659 MSNLSQAAK---------DAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYC 709
             :  : |          | | | | |   | |:  :  :|  |  ||| | ::
Db     815 CRDGEKCAHHSHHFSLDIDVGCTDLNE---DLGVWVIFKIKTQD-GHARLGNLEFLEE-- 868

Qy     710 GGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHN------------- 756
             |   | |    | |  |         :   |||   |::     |
Db     869 KPLVGEALARVKRAEK-KWRDKREKLEWETNIVYKEAKESVDALFVNSQYDQLQADTNIA 927

Qy     757 -LYESAKRLR-LRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVT 814
            :: : ||:  :||     |  :|  |: |: :| |       |: |   |:  : |
Db     928 MIHAADKRVHSIREAYLPELSVIPGVNAAIFEELEG------RIFTAFSLYDARNVI--- 978

Qy     815 LSKTGFFDAANHPLLQDTSRPTYKGFLD-ELLNN----------ISTINTDLDIEASGGY 863
             | | |:          |    || :| |  ||           : :: :: :    ||
Db     979 --KNGDFN-------NGLSCWNVKGHVDVEEQNNQRSVLVLPEWEAEVSQEVRVCPGRGY 1029

Qy     864 DDDLIARLLKL------GCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRME 917
               | |:         ||    ||   |   :||    ||   |    :  |    : |
Db    1030 ----ILRVTAYKEGYGEGCVTIHEIENNT-DELKFSNCVEEEIYPNNTVTCNDYTVNQEE 1084

Qy     918 QRMAY-----GHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEFGKVENGRSTTAMALTV 972
              ||     |:||   |              ||:   |:: |     :||
```

```
Db   1085 YGGAYTSRNRGYNEAPSV--------------PAD--YASVYEEKSYTDGRRENPCEFNR 1128

Qy    973 G----IPAAIGALLLLKNKVQTKGVIRPLQ--PEIYVPALEILESSGIKLVEKVE 1021
          |     |  :|             | : |:  ||     :|| |: |  :|: ||
Db   1129 GYRDYTPLPVGY------------VTKELEYFPETDKVWIEIGETEGTFIVDSVE 1171


RESULT  14
US-07-727-814B-2
; Sequence 2, Application US/07727814B
; Patent No. 5213969
;  GENERAL INFORMATION:
;    APPLICANT:  SCHUMACHER, Gunther
;    APPLICANT:  BURTSCHER, Helmut
;    APPLICANT:  MOLLERING, Hans
;    TITLE OF INVENTION:  CLONED N-METHYLHYDANTOINASE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Armstrong, Nikaido, Marmelstein, Kubovcik &
;      ADDRESSEE:  Murray
;      STREET:  1725 K Street, N.W., Suite 1000
;      CITY:  Washington D.C.
;      COUNTRY:  United States of America
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/727,814B
;      FILING DATE:  19910708
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 4021571.7
;      FILING DATE:  06-JUL-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murray, Robert B.
;      REGISTRATION NUMBER:  22,980
;      REFERENCE/DOCKET NUMBER:  911014
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2930
;      TELEFAX:  (202) 887-0357
;      TELEX:  440142
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1288 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-727-814B-2


  Query Match             2.1%;  Score 109.5;  DB 1;  Length 1288;
  Best Local Similarity   17.7%;  Pred. No. 0.64;
  Matches  104;  Conservative   93;  Mismatches  236;  Indels  153;  Gaps   22;

Qy    347 DMEGAGVVCLAVDILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAG 406
          ::: |||  :|| :| :  : |  |  | |::             | |     | :: :
Db    152 ELKTAGVQAIAVCLLHSYLNPEHEQRIGEIVNE------------EFP----EAYLSLSS 195

Qy    407 RLTPLYEYIPRMRNTMIDLAPAKTNPLPDKKYSTLVSLSGHL-FDKFLINEALDIIETAG 465
           : |||    |   | ::   |   |   :    |   : :| : : ::     :::::|
Db    196 EIVPLYREYERFSTTALN---AYVGPRVSRYLHRLQEQAENLGYQREIL-----LMQSSG 247

Qy    466 GSFHLVRCEVGQSTDDMS----------YSELEVGADDTATLDKIIDSLTSLANEHGGDH 515
          |   :        |  ||          ::  : | ::  |||             |
Db    248 GMVPIGEAAKRPVTLMMSGPVGGLIGGMWAAKQSGFENVVTLD-----------IGGTSA 296

Qy    516 DAGQEIELALKI-----GKVNEYETDV------TIDKGGPKILILGAGRVCRPAAEFLAS 564
          | |   :  |::      |: :::  |      ||  ||  |  : || | |   :   :
Db    297 DIGVAYQGELRMRHLLDTKIGDHQAMVPMVDIDTIGAGGGSIAYVDAGGVFRVGPQSAGA 356
```

```
Qy    565 YPDICTYG-----VDDHDADQI-------HVIVASLYQKD---AEETVDGIENTTATQLD 609
           |    ||         ||  :        ::  |    |   |   : |: :     ::
Db    357 VPGPVCYGRGGTEPTSTDAQVLLGRMRPDRILAGSGLDMDLDRARAAMQGLADKLGMSIE 416

Qy    610 VADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDA 669
           | :|:|     |::               |:   :::: |   |     : :  ||  |
Db    417 EAALGAL-----QIQKF-------------GMTQAIEQNSVRRGYDPRDFTLV--AAGGA 456

Qy    670 GVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWN 729
          |    ||:  :  : |::         |  | |               | |   : |::
Db    457 GALFACEIAAELEVPHVL-------VPAHPGIIAGIGLLATDEQYEFVATN--RFSFAFR 507

Qy    730 PAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLY 789
           |  :::                     ||  :| |  :| |   | :|
Db    508 DAAVIQAS--------------------YEQLERERNAQLDA---EEVP----------- 533

Qy    790 GISKEASTIYRATXRYEG--------FSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGFL 841
            ::    ::    ||||          |  || :     :|| |    :     :||
Db    534 --AERRKIVWLRDARYEGQGYEIRFVVPEGPVTTAWLDQAEAAFHDAHFEEYGHRFKGGT 591

Qy    842 DELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTV 887
           |::|        : :||    |:       : |  :|  :  : |
Db    592 VEVIN--------IRVEARAVMDELPTPEATQSGSLENALVETRPV 629


RESULT  15
US-08-258-614-2
; Sequence 2, Application US/08258614
; Patent No. 5432070
;  GENERAL INFORMATION:
;    APPLICANT:  SCHUMACHER, Gunther
;    APPLICANT:  BURTSCHER, Helmut
;    APPLICANT:  MOLLERING, Hans
;    TITLE OF INVENTION:  CLONED N-METHYLHYDANTOINASE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram
;      STREET:  655 Fifteenth Street N.W. Suite 330
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,614
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/021,856
;      FILING DATE:  24-FEB-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 40 21 571.7
;      FILING DATE:  06-JUL-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/727,814
;      FILING DATE:  08-JUL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kitts, Monica C.
;      REGISTRATION NUMBER:  36,105
;      REFERENCE/DOCKET NUMBER:  P564-3007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)638-5000
;      TELEFAX:  (202)638-4810
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1288 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

```
US-08-258-614-2

  Query Match             2.1%;  Score 109.5;  DB 1;  Length 1288;
  Best Local Similarity  17.7%;  Pred. No. 0.64;
  Matches  104;  Conservative   93;  Mismatches  236;  Indels  153;  Gaps   22;

Qy     347 DMEGAGVVCLAVDILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAG 406
           ::: |||  :|| :| :  : |  |  | |::             | |     | :: :
Db     152 ELKTAGVQAIAVCLLHSYLNPEHEQRIGEIVNE------------EFP----EAYLSLSS 195

Qy     407 RLTPLYEYIPRMRNTMIDLAPAKTNPLPDKKYSTLVSLSGHL-FDKFLINEALDIIETAG 465
            : |||    |   | ::   |   |   :    |   : :| : : ::     :::::|
Db     196 EIVPLYREYERFSTTALN---AYVGPRVSRYLHRLQEQAENLGYQREIL-----LMQSSG 247

Qy     466 GSFHLVRCEVGQSTDDMS----------YSELEVGADDTATLDKIIDSLTSLANEHGGDH 515
           |   :        |  ||          ::  : | ::  |||             |
Db     248 GMVPIGEAAKRPVTLMMSGPVGGLIGGMWAAKQSGFENVVTLD-----------IGGTSA 296

Qy     516 DAGQEIELALKI-----GKVNEYETDV------TIDKGGPKILILGAGRVCRPAAEFLAS 564
           | |   :  |::      |: :::  |      ||  ||  |  : || | |   :   :
Db     297 DIGVAYQGELRMRHLLDTKIGDHQAMVPMVDIDTIGAGGGSIAYVDAGGVFRVGPQSAGA 356

Qy     565 YPDICTYG-----VDDHDADQI-------HVIVASLYQKD---AEETVDGIENTTATQLD 609
            |    ||         ||  :        ::  |    |   |   : |: :     ::
Db     357 VPGPVCYGRGGTEPTSTDAQVLLGRMRPDRILAGSGLDMDLDRARAAMQGLADKLGMSIE 416

Qy     610 VADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDA 669
            | :|:|     |::               |:   :::: |   |     : :  ||  |
Db     417 EAALGAL-----QIQKF-------------GMTQAIEQNSVRRGYDPRDFTLV--AAGGA 456

Qy     670 GVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWN 729
           |    ||:  :  : |::         |  | |               | |   : |::
Db     457 GALFACEIAAELEVPHVL-------VPAHPGIIAGIGLLATDEQYEFVATN--RFSFAFR 507

Qy     730 PAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLY 789
            |  :::                     ||  :| |  :| |   | :|
Db     508 DAAVIQAS--------------------YEQLERERNAQLDA---EEVP----------- 533

Qy     790 GISKEASTIYRATXRYEG--------FSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGFL 841
             ::    ::    ||||          |  || :     :|| |    :     :||
Db     534 --AERRKIVWLRDARYEGQGYEIRFVVPEGPVTTAWLDQAEAAFHDAHFEEYGHRFKGGT 591

Qy     842 DELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTV 887
            |::|        : :||    |:       : |. :|  :  : |
Db     592 VEVIN--------IRVEARAVMDELPTPEATQSGSLENALVETRPV 629


Search completed: May 31, 2002, 15:48:06
Job time: 163 sec
```

```
                        GenCore version 4.5
                 Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          May 31, 2002, 15:44:38 ; Search time 39.13 Seconds
                                           (without alignments)
                                           2901.033 Million cell updates/sec

Title:           US-09-049-304A-122
Perfect score:   5286
Sequence:        1 CARLLLGGGKNGPRVNRIIV..........VPALEILESSGIKLVEKVET 1022

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters:      747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_032802:*
              1:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
              2:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
              3:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
              4:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
              5:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
              6:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
              7:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
              8:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
              9:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
             10:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
             11:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
             12:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
             13:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
             14:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
             15:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
             16:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
             17:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
             18:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
             19:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
             20:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
             21:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
             22:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

   Pred. No. is the number of results predicted by chance to have a
   score greater than or equal to the score of the result being printed,
   and is derived by analysis of the total score distribution.

                              SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                      Description
-----------------------------------------------------------------------
    1    5284   100.0   1022 20  AAW87761                Maize lysine ketog
    2    3416    64.6   1064 20  AAW87759                Arabidopsis lysine
    3    3411    64.5   1064 21  AAG31238                Arabidopsis thalia
    4    3316    62.7    640 20  AAW87762                Maize lysine ketog
    5  3190.5    60.4    969 21  AAG31239                Arabidopsis thalia
    6  3089.5    58.4    948 21  AAG31240                Arabidopsis thalia
    7  2498.5    47.3    784 20  AAW87760                Soybean lysine ket
    8    1480    28.0    482 20  AAW87766                Arabidopsis saccha
    9  1466.5    27.7    928 22  AAE05075                Drosophila melanog
   10  1466.5    27.7    972 22  ABB63015                Drosophila melanog
   11     996    18.8    239 20  AAW87763                Rice lysine ketogl
   12     360     6.8     83 20  AAW87765                Wheat lysine ketog
   13     270     5.1     74 19  AAW60533                Protein homologous
   14     270     5.1     74 20  AAW87758                Arabidopsis lysine
   15   263.5     5.0    123 19  AAW60532                Protein homologous
   16   263.5     5.0    123 20  AAW87757                Arabidopsis lysine
   17   181.5     3.4    145 22  AAB94624                Human protein sequ
   18     179     3.4     42 20  AAW87764                Rice lysine ketogl
   19   144.5     2.7    382 17  AAW00496                Candida albicans s
   20   130.5     2.5    605 11  AAR03923                Bacillus megateriu
   21   125.5     2.4   1104 21  AAB23856                Haemophilus influe
   22   125.5     2.4   1104 21  AAB23859                Haemophilus influe
   23     120     2.3    491 15  AAR60342                Partial human Lipi
   24     120     2.3    493 20  AAY49556                Human cholesteryl
   25     120     2.3    493 20  AAY49568                Human cholesteryl
   26     120     2.3    493 22  AAG65636                Human cholesterol
   27   119.5     2.3   1349 22  AAM79343                Human protein SEQ
   28   119.5     2.3   1349 22  AAM40485                Human polypeptide
   29   119.5     2.3   1349 22  AAM40486                Human polypeptide
   30     119     2.3    774 20  AAY37454                Protein involved i
   31   117.5     2.2   1194 22  AAM78359                Human protein SEQ
   32   117.5     2.2   1194 22  AAM38699                Human polypeptide
   33     117     2.2   1118 22  AAB48264                Rice magnesium che
   34     116     2.2    632 22  AAU35319                Enterococcus faeca
   35     116     2.2    923 22  ABB59510                Drosophila melanog
   36   114.5     2.2   3177 22  AAM40303                Human polypeptide
   37   112.5     2.1   1243 20  AAY35286                C. pneumoniae prot
   38     112     2.1    476 17  AAW06127                Human cholesteryl
   39     112     2.1    476 18  AAW46446                Human mature chole
   40     112     2.1    476 20  AAY02466                Human cholesteryl
   41   111.5     2.1    531 22  ABB60201                Drosophila melanog
   42   111.5     2.1   1023 16  AAR76991                LhaA (low homology
   43   111.5     2.1   1421 21  AAG39649                Arabidopsis thalia
   44   111.5     2.1   1499 21  AAG39648                Arabidopsis thalia
   45   111.5     2.1   1655 21  AAG39647                Arabidopsis thalia


                              ALIGNMENTS


RESULT   1
AAW87761
ID   AAW87761 standard; Protein; 1022 AA.
XX
AC   AAW87761;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Maize lysine ketoglutarate reductase.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; maize.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 803
FT                   /note= "encoded by YTT"
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   N-PSDB; AAV99563.
XX
```

```
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 2; Page 189-191; 231pp; English.
XX
CC   This is the amino acid sequence of a maize near full-length
CC   lysine ketoglutarate reductase (LKR) or saccharopine dehydrogenase
CC   (SDH).  It was deduced from cDNA (see AAV99563) obtained from
CC   developing seed mRNA by PCR and RACE.  Isolated nucleic acids
CC   comprising sequences encoding all or part of plant LKR enzymes are
CC   new.  Also claimed are: (1) a chimeric gene comprising the
CC   fragment (or a subfragment) operably linked to a seed specific
CC   regulator, where the chimeric gene reduces LKR activity in plant
CC   seeds transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence   1022 AA;

  Query Match             100.0%;  Score 5284;  DB 20;  Length 1022;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1022;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 carlllgggkngprvnriivqpstrrihhdaqyedagceisedlsecgliigikqpklqm 60

Qy       61 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 ilsdrayaffshthkaqkenmplldkileervslfdyelivgddgkrslafgkfagragl 120

Qy      121 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 idflhglgqrylslgystpflslgqshmypslaaakaavivvaeeiatfglpsgicpivf 180

Qy      181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRD 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 vftgvgnvsqgaqeifkllphtfvdaeklpeifqarnlskqsqstkrvfqlygcvvtsrd 240

Qy      241 IVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQL 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 ivshkdptrqfdkgdyyahpehytpvfheriapyasvivncmywekrfppllnmdqlqql 300

Qy      301 METGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDI 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 metgcplvgvcditcdiggsiefinkstsierpffrydpsknsyhddmegagvvclavdi 360

Qy      361 LPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMRN 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 lptefskeasqhfgnilsrlvaslasvkqpaelpsylrraciahagrltplyeyiprmrn 420

Qy      421 TMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQSTD 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 tmidlapaktnplpdkkystlvslsghlfdkflinealdiietaggsfhlvrcevgqstd 480

Qy      481 DMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYETDVTID 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 dmsyselevgaddtatldkiidsltslanehggdhdagqeielalkigkvneyetdvtid 540

Qy      541 KGGPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGI 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 kggpkililgagrvcrpaaeflasypdictygvddhdadqihvivaslyqkdaeetvdgi 600

Qy      601 ENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMS 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 enttatqldvadigslsdlvsqvevvisllpasfhaaiagvcielkkhmvtasyvdesms 660

Qy      661 NLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANN 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 nlsqaakdagvtilcemgldpgidhlmsmkmideaharkgkikaftsycgglpspaaann 720

Qy      721 PLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNR 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 playkfswnpagalrsgknpavykflgetihvdghnlyesakrlrlrelpafalehlpnr 780

Qy      781 NSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGF 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 nsliygdlygiskeastiyratxryegfseimvtlsktgffdaanhpllqdtsrptykgf 840

Qy      841 LDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHEETQ 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 ldellnnistintdldieasggydddliarllklgccknkeiavktvktikflglheetq 900

Qy      901 IPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEFGKVE 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 ipkgcsspfdvicqrmeqrmayghneqdmvllhheveveypdgqpaekhqatllefgkve 960

Qy      961 NGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSGIKLVEKV 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 ngrsttamaltvgipaaigallllknkvqtkgvirplqpeiyvpaleilessgiklvekv 1020

Qy     1021 ET 1022
            ||
Db     1021 et 1022


RESULT   2
AAW87759
ID   AAW87759 standard; Protein; 1064 AA.
XX
AC   AAW87759;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Arabidopsis lysine ketoglutarate reductase.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed.
XX
OS   Arabidopsis thaliana cv. Landsberg erecta.
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   N-PSDB; AAV99559.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
```

```
PT   transformed plants
XX
PS   Claim 2; Page 175-178; 231pp; English.
XX
CC   This is the amino acid sequence of Arabidopsis thaliana
CC   lysine ketoglutarate reductase (LKR)/saccharopine dehydrogenase
CC   (SDH), as deduced from a gene (see AAV99599) isolated from a CD4-8
CC   Landsberg erecta genomic library.  The protein lacks an N-terminal
CC   targetting sequence suggesting that the lysine degradative
CC   pathway in Arabidopsis is located in the plant cell cytosol.
CC   Isolated nucleic acid fragments comprising sequences encoding all
CC   or part of plant LKR polypeptides are new.  Also claimed are: (1) a
CC   chimeric gene comprising the fragment (or a subfragment) operably
CC   linked to a suitable seed specific regulator, where the chimeric
CC   gene reduces LKR activity in plant seeds transformed with it; (2)
CC   plant cells and seeds in which LKR activity is reduced due to a
CC   mutation in the gene encoding LKR or transformation with the
CC   chimeric gene; (3) a nucleic acid fragment comprising: (i) chimeric
CC   gene above, and (ii) a second chimeric gene, in which a nucleic
CC   acid fragment encoding dihydrodipicolinic acid synthase (DHDPS)
CC   substantially insensitive to lysine inhibition is operably linked
CC   to a plant chloroplast transit sequence and to a plant
CC   seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence   1064 AA;


  Query Match             64.6%;  Score 3416;  DB 20;  Length 1064;
  Best Local Similarity   63.6%;  Pred. No. 5.2e-295;
  Matches  668;  Conservative  152;  Mismatches  178;  Indels   52;  Gaps   13;

Qy        1 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 60
            ||| || |||:   ::||:|||| :|||||| ||  |||||:|||:||||:|||||:|:|
Db       38 car-llhggkdrtgisrivvqpsakrihhdalyehvgceisddlsdcglilgikqpelem 96

Qy       61 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 120
            || :|||||||||||||||||||||||| |||:| |||||||| ||| |||||:||||||
Db       97 ilperayaffshthkaqkenmplldkilservtlcdyelivgdhgkrllafgkyagragl 156

Qy      121 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
            :||||||||||||||||||||||| |:|| |||||||||| | ||||: ||| ||||:||
Db      157 vdflhglgqrylslgystpflslgasymysslaaakaavisvgeeiasqglplgicplvf 216

Qy      181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIF-QARNLSKQSQSTKRVFQLYGCVVTSR 239
            |||| |||| ||||||||||||||:  ||||:| : : :|:   |||||:|:|||::||:
Db      217 vftgtgnvslgaqeifkllphtfvepsklpelfvkdkgisqngistkrvyqvygciitsq 276

Qy      240 DIVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQ 299
            |:| ||||:: ||| ||||||||| |||||:|:|| ||:||||||||||| ||:  ||| 
Db      277 dmvehkdpsksfdkadyyahpehynpvfhekispytsvlvncmywekrfpcllstkqlqd 336

Qy      300 LMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVD 359
            | : | ||||:|||||||||||||:|::| |: ||||::|| |||:|||:| ||:|:|||
Db      337 ltkkglplvgicditcdiggsiefvnratlidspffrfnpsnnsyyddmdgdgvlcmavd 396

Qy      360 ILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMR 419
            ||||||:||||||||:|||  | ||||: : ::||::|:|||||:: | || |||||||||
Db      397 ilptefakeasqhfgdilsgfvgslasmteisdlpahlkracisyrgeltslyeyiprmr 456

Qy      420 NTMIDLAPAKTNP------------LPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGS 467
                     |:||               : :: |||||||||||||||||||:|| ||||
Db      457 ---------ksnpeeaqdniiangvssqrtfnilvslsghlfdkflinealdmieaaggs 507

Qy      468 FHLVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLAN---EHGGDHDAGQEIELA 524
            ||| :||:||| |  |||||||||||   ||:|||||| |||   ::   |    :|  :
Db      508 fhlakcelgqsadaesyselevgaddkrvldqiidsltrlanpnedyisphreanki--s 565

Qy      525 LKIGKVNEYETDVTIDKGGPK------ILILGAGRVCRPAAEFLASYPDICT-------Y 571
            |||||| : | ::   |  |:      :|||||||||||||:||||   | :       :
Db      566 lkigkvqq-enei---kekpemtkksgvlilgagrvcrpaadflasvrtissqqwyktyf 621

Qy      572 GVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLP 631
            | |  :   :|||||||| |||:|||:|| :  | :|||:|  ||   ||||:||:||||
Db      622 gadseektdvhvivaslylkdaketvegisdveavrldvsdsesllkyvsqvdvvlsllp 681

Qy      632 ASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKM 691
            || || :|  |||||||:|||||||:  | | : || ||:||| ||||||||||:|:|||
Db      682 aschavvaktcielkkhlvtasyvddetsmlhekaksagitilgemgldpgidhmmamkm 741

Qy      692 IDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIH 751
            |::|| :|||:|:||||||||||||||||||||||||||||||:|:|:||| ||  |: ||
Db      742 indahikkgkvksftsycgglpspaaannplaykfswnpagairagqnpakyksngdiih 801

Qy      752 VDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEI 811
            ||| |||:|| | |:  |||||||  |||:||:||: |||  ||:||:| | |||||| |
Db      802 vdgknlydsaarfrvpnlpafalecfpnrdslvygehygieseattifrgtlryegfsmi 861

Qy      812 MVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIARL 871
            | |||| ||||:  : :|    | |:   |  :||     : | : |   | ::::  |:
Db      862 matlsklgffdseanqvlstgkritfgallsnilnk----dadneseplag-eeeiskri 916

Qy      872 LKLGCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVL 931
            :|||   :|| | |  ||| ||| :|| ::|  | | ||  |  ||:::||  |||||||
Db      917 iklg--hsketaakaaktivflgfneerevpslcksvfdatcylmeeklaysgneqdmvl 974

Qy      932 LHHEVEVEYPDGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTK 991
            ||||||||: : :  ||| ||||||| ::||::||||| |||||||||||:|:::|::|:
Db      975 lhheveveflesktriekhtatllefgdikngqtttamaktvgipaaigalvliedkiktr 1034

Qy      992 GVIRPLQPEIYVPALEILESSGIKLVEKVE 1021
            ||:|||: |:|:|||:||:: ||||:|| |
Db     1035 gvlrpleaevylpaldilqaygiklmekae 1064


RESULT   3
AAG31238
ID   AAG31238 standard; Protein; 1064 AA.
XX
AC   AAG31238;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 37481.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
```

```
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
```

```
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
```

```
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             64.5%;  Score 3411;  DB 21;  Length 1064;
  Best Local Similarity   63.6%;  Pred. No. 1.5e-294;
  Matches  668;  Conservative  151;  Mismatches  179;  Indels   52;  Gaps   13;

Qy      1 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 60
          ||| || |||:   ::||:|||| :|||||| ||| |||||:|||:||||:|||||:|:|
Db     38 car-llhggkdrtgisrivvqpsakrihhdalyedvgceisddlsdcglilgikqpelem 96

Qy     61 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 120
          || :|||||||||||||||||||||||| |||:| |||||||| ||| |||||:||||||
Db     97 ilperayaffshthkaqkenmplldkilservtlcdyelivgdhgkrllafgkyagragl 156

Qy    121 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
          :||||||||| | ||||||||||| |:|| |||||||||| | ||||: ||| ||||:||
Db    157 vdflhglgqrklilgystpflslgasymysslaaakaavisvgeeiasqglplgicplvf 216

Qy    181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIF-QARNLSKQSQSTKRVFQLYGCVVTSR 239
          |||| |||| ||||||||||||||:  ||||:| : : :|:   |||||:|:|||::||:
Db    217 vftgtgnvslgaqeifkllphtfvepsklpelfvkdkgisqngistkrvyqvygciitsq 276

Qy    240 DIVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQ 299
          |:| ||||:: ||| ||||||||| |||||:|:|| ||:||||||||||| ||:  ||| 
Db    277 dmvehkdpsksfdkadyyahpehynpvfhekispytsvlvncmywekrfpcllstkqlqd 336

Qy    300 LMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVD 359
          | : | ||||:|||||||||||||:|::| |: ||||::|| |||:|||:| ||:|:|||
Db    337 ltkkglplvgicditcdiggsiefvnratlidspffrfnpsnnsyyddmdgdgvlcmavd 396

Qy    360 ILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMR 419
          ||||||:||||||||:|||  | ||||: : ::||::|:||||:: | || |||||||||
Db    397 ilptefakeasqhfgdilsgfvgslasmteisdlpahlkracisyrgeltslyeyiprmr 456

Qy    420 NTMIDLAPAKTNP------------LPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGS 467
                   |:||               : :: |||||||||||||||||||:|| ||||
Db    457 ---------ksnpeeaqdniiangvssqrtfnilvslsghlfdkflinealdmieaaggs 507

Qy    468 FHLVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLAN---EHGGDHDAGQEIELA 524
          ||| :||:||| |  |||||||||||   ||:|||||| |||   ::   |    :|  :
Db    508 fhlakcelgqsadaesyselevgaddkrvldqiidsltrlanpnedyisphreanki--s 565

Qy    525 LKIGKVNEYETDVTIDKGGPK------ILILGAGRVCRPAAEFLASYPDICT-------Y 571
          |||||| : | ::   |  |:      :|||||||||||||:||||   | :       :
Db    566 lkigkvqq-enei---kekpemtkksgvlilgagrvcrpaadflasvrtissqqwyktyf 621

Qy    572 GVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLP 631
          | |  :   :|||||||| |||:|||:|| :  | :|||:|  ||   ||||:||:||||
Db    622 gadseektdvhvivaslylkdaketvegisdveavrldvsdsesllkyvsqvdvvlsllp 681

Qy    632 ASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKM 691
          || || :|  |||||||:|||||||:  | | : || ||:||| ||||||||||:|:|||
Db    682 aschavvaktcielkkhlvtasyvddetsmlhekaksagitilgemgldpgidhmmamkm 741

Qy    692 IDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIH 751
          |::|| :|||:|:|||||||||||||||||||||||||||||:|:|:||| ||  |: ||
Db    742 indahikkgkvksftsycgglpspaaannplaykfswnpagairagqnpakyksngdiih 801

Qy    752 VDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEI 811
          ||| |||:|| | |:  |||||||  |||:||:||: |||  ||:||:| | |||||| |
Db    802 vdgknlydsaarfrvpnlpafalecfpnrdslvygehygieseattifrgtlryegfsmi 861

Qy    812 MVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIARL 871
          | |||| ||||:  : :|    | |:   |  :||     : | : |   | ::::  |:
Db    862 matlsklgffdseanqvlstgkritfgallsnilnk----dadneseplag-eeeiskri 916

Qy    872 LKLGCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVL 931
          :|||   :|| | |  ||| ||| :|| ::|  | | ||  |  ||:::||  |||||||
Db    917 iklg--hsketaakaaktivflgfneerevpslcksvfdatcylmeeklaysgneqdmvl 974

Qy    932 LHHEVEVEYPDGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTK 991
          ||||||||: : :  ||| ||||||| ::||::||||| |||||||||||||:::|::|:
Db    975 lhheveveflesskriekhtatllefgdikngqtttamaktvgipaaigallliedkiktr 1034

Qy    992 GVIRPLQPEIYVPALEILESSGIKLVEKVE 1021
          ||:|||: |:|:|||:||:: ||||:|| |
Db   1035 gvlrpleaevylpaldilqaygiklmekae 1064


RESULT   4
AAW87762
ID   AAW87762 standard; Protein; 640 AA.
XX
AC   AAW87762;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Maize lysine ketoglutarate reductase partial sequence.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; maize.
XX
OS   Zea mays.
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   N-PSDB; AAV99564.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 2; Page 193-195; 231pp; English.
XX
CC   This is the amino acid sequence of a maize lysine ketoglutarate
CC   reductase (LKR) or saccharopine dehydrogenase (SDH) partial
CC   polypeptide.  It was deduced from cDNA (see AAV99564) obtained from
CC   developing seed mRNA by PCR and RACE.  Isolated nucleic acids
CC   comprising sequences encoding all or part of plant LKR enzymes are
CC   new.  Also claimed are: (1) a chimeric gene comprising the
CC   fragment (or a subfragment) operably linked to a seed specific
CC   regulator, where the chimeric gene reduces LKR activity in plant
```

```
CC   seeds transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence   640 AA;

  Query Match             62.7%;  Score 3316;  DB 20;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-286;
  Matches  640;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 carlllgggkngprvnriivqpstrrihhdaqyedagceisedlsecgliigikqpklqm 60

Qy         61 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ilsdrayaffshthkaqkenmplldkileervslfdyelivgddgkrslafgkfagragl 120

Qy        121 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 idflhglgqrylslgystpflslgqshmypslaaakaavivvaeeiatfglpsgicpivf 180

Qy        181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRD 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 vftgvgnvsqgaqeifkllphtfvdaeklpeifqarnlskqsqstkrvfqlygcvvtsrd 240

Qy        241 IVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ivshkdptrqfdkgdyyahpehytpvfheriapyasvivncmywekrfppllnmdqlqql 300

Qy        301 METGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDI 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 metgcplvgvcditcdiggsiefinkstsierpffrydpsknsyhddmegagvvclavdi 360

Qy        361 LPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMRN 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 lptefskeasqhfgnilsrlvaslasvkqpaelpsylrraciahagrltplyeyiprmrn 420

Qy        421 TMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQSTD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 tmidlapaktnplpdkkystlvslsghlfdkflinealdiietaggsfhlvrcevgqstd 480

Qy        481 DMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYETDVTID 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 dmsyselevgaddtatldkiidsltslanehggdhdagqeielalkigkvneyetdvtid 540

Qy        541 KGGPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 kggpkililgagrvcrpaaeflasypdictygvddhdadqihvivaslyqkdaeetvdgi 600

Qy        601 ENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAG 640
              ||||||||||||||||||||||||||||||||||||||||
Db        601 enttatqldvadigslsdlvsqvevvisllpasfhaaiag 640

RESULT 5
AAG31239
ID   AAG31239 standard; Protein; 969 AA.
XX
AC   AAG31239;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 37482.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
```

```
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
```

```
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             60.4%;  Score 3190.5;  DB 21;  Length 969;
  Best Local Similarity   63.3%;  Pred. No. 5.8e-275;
  Matches  627;  Conservative  141;  Mismatches  172;  Indels   51;  Gaps   12;

Qy         60 MILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAG 119
              ||| :|||||||||||||||||||||||| |||:| |||||||| ||| |||||:|||||
Db          1 milperayaffshthkaqkenmplldkilservtlcdyelivgdhgkrllafgkyagrag 60

Qy        120 LIDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIV 179
              |:||||||||| | ||||||||||| |:|| |||||||||| | ||||: ||| ||||:|
Db         61 lvdflhglgqrklilgystpflslgasymysslaaakaavisvgeeiasqglplgicplv 120

Qy        180 FVFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIF-QARNLSKQSQSTKRVFQLYGCVVTS 238
              ||||| |||| ||||||||||||||:  ||||:| : : :|:   |||||:|:|||::||
Db        121 fvftgtgnvslgaqeifkllphtfvepsklpelfvkdkgisqngistkrvyqvygciits 180

Qy        239 RDIVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQ 298
              :|:| ||||:: ||| ||||||||| |||||:|:|| ||:||||||||||| ||:  |||
Db        181 qdmvehkdpsksfdkadyyahpehynpvfhekispytsvlvncmywekrfpcllstkqlq 240
```

```
Qy   299 QLMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAV 358
          | : | ||||:|||||||||||||:|::| |: ||||::|| |||:|||:| ||:|:||
Db   241 dltkkglplvgicditcdiggsiefvnratlidspffrfnpsnnsyyddmdgdgvlcmav 300

Qy   359 DILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRM 418
         |||||||:||||||||:|||  | ||||: : ::||::|:||||:: | || ||||||||
Db   301 dilptefakeasqhfgdilsgfvgslasmteisdlpahlkracisyrgeltslyeyiprm 360

Qy   419 RNTMIDLAPAKTNP------------LPDKKYSTLVSLSGHLFDKFLINEALDIIETAGG 466
         |         |:||               : :: |||||||||||||||||||:|| |||
Db   361 r---------ksnpeeaqdniiangvssqrtfnilvslsghlfdkflinealdmieaagg 411

Qy   467 SFHLVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLAN---EHGGDHDAGQEIEL 523
         |||| :||:||| |  |||||||||||   ||:|||||| |||   ::   |    :|
Db   412 sfhlakcelgqsadaesyselevgaddkrvldqiidsltrlanpnedyisphreanki-- 469

Qy   524 ALKIGKVNEYETDVTIDKGGPK------ILILGAGRVCRPAAEFLASYPDICT------- 570
         :||||||  : | ::   |  |:      ::||||||||||||:||||   | :
Db   470 slkigkvqq-enei---kekpemtkksgvlilgagrvcrpaadflasvrtissqqwykty 525

Qy   571 YGVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLL 630
         :| |  :   :|||||||| |||:|||:|| :  | :|||:|  ||   ||||:||:|||
Db   526 fgadseektdvhvivaslylkdaketvegisdveavrldvsdsesllkyvsqvdvvlsll 585

Qy   631 PASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMK 690
         ||| || :|  |||||||:|||||||:  | | : || ||:||| ||||||||||:|:||
Db   586 paschavvaktcielkkhlvtasyvddetsmlhekaksagitilgemgldpgidhmmamk 645

Qy   691 MIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETI 750
         ||::|| :|||:|:|||||||||||||||||||||||||||||:|:|:||| ||  |: |
Db   646 mindahikkgkvksftsycgglpspaaannplaykfswnpagairagqnpakyksngdii 705

Qy   751 HVDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSE 810
         |||| |||:|| | |:  |||||||  |||:||:||: |||  ||:||:| | ||||||
Db   706 hvdgknlydsaarfrvpnlpafalecfpnrdslvygehygieseattifrgtlryegfsm 765

Qy   811 IMVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIAR 870
         || |||| ||||:  : :|    | |:   |  :||     : | : |   | ::::  |
Db   766 imatlsklgffdseanqvlstgkritfgallsnilnk----dadneseplag-eeeiskr 820

Qy   871 LLKLGCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMV 930
         ::|||   :|| | |  ||| ||| :|| ::|  | | ||  |  ||:::||  ||||||
Db   821 iiklg--hsketaakaaktivflgfneerevpslcksvfdatcylmeeklaysgneqdmv 878

Qy   931 LLHHEVEVEYPDGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQT 990
         |||||||||: : :  ||| ||||||| ::||::||||| |||||||||||||:::|::|
Db   879 llhhevevefleskriekhtatllefgdikngqtttamaktvgipaaigallliedkikt 938

Qy   991 KGVIRPLQPEIYVPALEILESSGIKLVEKVE 1021
         :||:|||: |:|:|||:||:: ||||:|| |
Db   939 rgvlrpleaevylpaldilqaygiklmekae 969

RESULT    6
AAG31240
ID   AAG31240 standard; Protein; 948 AA.
XX
AC   AAG31240;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 37483.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
```

```
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             58.4%;  Score 3089.5;  DB 21;  Length 948;
  Best Local Similarity   62.7%;  Pred. No. 5.7e-266;
  Matches  608;  Conservative  140;  Mismatches  171;  Indels   51;  Gaps   12;

Qy         81 MPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRYLSLGYSTPF 140
              |||||||| |||:| |||||||| ||| |||||:||||||:||||||||| | |||||||
Db          1 mplldkilservtlcdyelivgdhgkrllafgkyagraglvdflhglgqrklilgystpf 60

Qy        141 LSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVSQGAQEIFKLLP 200
              |||| |:|| |||||||||| | ||||: ||| ||||:|||||| |||| ||||||||||
Db         61 lslgasymysslaaakaavisvgeeiasqglplgicplvfvftgtgnvslgaqeifkllp 120

Qy        201 HTFVDAEKLPEIF-QARNLSKQSQSTKRVFQLYGCVVTSRDIVSHKDPTRQFDKGDYYAH 259
              ||||:  ||||:| : : :|:   |||||:|:|||::||:|:| ||||:: ||| |||||
Db        121 htfvepsklpelfvkdkgisqngistkrvyqvygciitsqdmvehkdpsksfdkadyyah 180

Qy        260 PEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLMETGCPLVGVCDITCDIGG 319
              |||| |||||:|:|| ||:||||||||||| ||:  ||| | : | ||||:|||||||||
Db        181 pehynpvfhekispytsvlvncmywekrfpcllstkqlqdltkkglplvgicditcdigg 240

Qy        320 SIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDILPTEFSKEASQHFGNILSR 379
              ||||:|::| |: ||||::|| |||:|||:| ||:|:|||||||||:||||||||:|||
Db        241 siefvnratlidspffrfnpsnnsyyddmdgdgvlcmavdilptefakeasqhfgdilsg 300

Qy        380 LVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMRNTMIDLAPAKTNP------- 432
               | ||||: : ::||::|:||||::  | || |||||||||         |:||
Db        301 fvgslasmteisdlpahlkracisyrgeltslyeyiprmr---------ksnpeeaqdni 351

Qy        433 -----LPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQSTDDMSYSEL 487
                      : :: |||||||||||||||||||:|| ||||||| :||:||| |  |||||
Db        352 iangvssqrtfnilvslsghlfdkflinealdmieaaggsfhlakcelgqsadaesysel 411

Qy        488 EVGADDTATLDKIIDSLTSLAN---EHGGDHDAGQEIELALKIGKVNEYETDVTIDKGGP 544
              ||||||   ||:|||||| |||   ::   |    :|  :|||||| : | ::   |  |
Db        412 evgaddkrvldqiidsltrlanpnedyisphreanki--slkigkvqq-enei---kekp 465
```

```
Qy      545 K------ILILGAGRVCRPAAEFLASYPDICT-------YGVDDHDADQIHVIVASLYQK 591
            :      :|||||||||||||:||||   | :       :| |  :   :|||||||| |
Db      466 emtkksgvlilgagrvcrpaadflasvrtissqqwyktyfgadseektdvhvivaslylk 525

Qy      592 DAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVT 651
            ||:|||:|| :  | :|||:|  ||   ||||:||:|||||| || :|  |||||||:||
Db      526 daketvegisdveavrldvsdsesllkyvsqvdvvlsllpaschavvaktcielkkhlvt 585

Qy      652 ASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGG 711
            |||||:  | | : || ||:||| ||||||||||:|:||||::|| :|||:|:|||||||
Db      586 asyvddetsmlhekaksagitilgemgldpgidhmmamkmindahikkgkvksftsycgg 645

Qy      712 LPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPA 771
            ||||||||||||||||||||||:|:|:||| ||  |: ||||| |||:|| | |:  |||
Db      646 lpspaaannplaykfswnpagairagqnpakyksngdiihvdgknlydsaarfrvpnlpa 705

Qy      772 FALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQD 831
            ||||  |||:||:||: |||  ||:||:| | |||||| || |||| ||||:  : :|
Db      706 falecfpnrdslvygehygieseattifrgtlryegfsmimatlsklgffdseanqvlst 765

Qy      832 TSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIK 891
              | |:   |  :||     : | : |   | ::::  |::|||   :|| | |  |||
Db      766 gkritfgallsnilnk----dadneseplag-eeeiskriiklg--hsketaakaaktiv 818

Qy      892 FLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQA 951
            ||| :|| ::|  | | ||  |  ||:::||  |||||||||||||||: : :  ||| |
Db      819 flgfneerevpslcksvfdatcylmeeklaysgneqdmvllhhevevefleskriekhta 878

Qy      952 TLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILES 1011
            |||||| ::||::||||| |||||||||||||:::|::|:||:|||:: |:|:|||:||::
Db      879 tllefgdikngqtttamaktvgipaaigallliedkiktrgvlrpleaevylpaldilqa 938

Qy     1012 SGIKLVEKVE 1021
             ||||:|| |
Db      939 ygiklmekae 948

RESULT   7
AAW87760
ID   AAW87760 standard; Protein; 784 AA.
XX
AC   AAW87760;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Soybean lysine ketoglutarate reductase.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; soybean.
XX
OS   Glycine max.
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   N-PSDB; AAV99562.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 2; Page 186-188; 231pp; English.
XX
CC   This is the amino acid sequence of a soybean near full-length
CC   lysine ketoglutarate reductase (LKR) or saccharopine dehydrogenase
CC   (SDH).  It was deduced from cDNA (see AAV99562) obtained from
CC   developing seed mRNA by PCR and RACE.  Isolated nucleic acids
CC   comprising sequences encoding all or part of plant LKR enzymes are
CC   new.  Also claimed are: (1) a chimeric gene comprising the
CC   fragment (or a subfragment) operably linked to a seed specific
CC   regulator, where the chimeric gene reduces LKR activity in plant
CC   seeds transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence   784 AA;

  Query Match             47.3%;  Score 2498.5;  DB 20;  Length 784;
  Best Local Similarity   63.1%;  Pred. No. 2.1e-213;
  Matches  497;  Conservative  111;  Mismatches  163;  Indels   17;  Gaps    9;

Qy      245 KDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLMETG 304
            ||    ||| |||:||||| | |||:||||||||||||||||||| | :  |:| ||  |
Db        3 kdhvivfdkadyyshpehynptfhekiapyasvivncmywekrfpqlpsykqmqdlmgrg 62

Qy      305 CPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDILPTE 364
             ||||: ||||||||||||:|:|: |||: |||||||  ||||||||| ||:||||||||||
Db       63 splvgiaditcdiggsiefvnrgtsidspffrydpltnsyhddmegngviclavdilpte 122

Qy      365 FSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMRNT-MI 423
            |:||||||||||||: | :|||     :||::|||||||| | || ||:||||||::
Db      123 fakeasqhfgnilsqfvvnlasatditklpahlrraciahkgvltslydyiprmrssdse 182

Qy      424 DLAPAKTNPLPDK-KYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQSTDDM 482
            :::    | | :| ||:  |||||||||:||||||||||| |||||||| | |||| : :
Db      183 evsenaenslsnkrkynisvslsghlfdqflinealdiieaaggsfhlvnchvgqsieav 242

Qy      483 SYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQE-IELALKIGKVNEYETDVTID- 540
            |:||||||||: | ||:||||||::|:    |  : |:  :::||:||| |   :   |
Db      243 sfselevgadnravldqiidsltaiasptehdrfsnqdsskislklgkveengiekesdp 302

Qy      541 KGGPKILILGAGRVCRPAAEFLASY--PDIC----TYGVDDHDAD-QIHVIVASLYQKDA 593
            :    :|||||||||:|||| |:|:  |       |   || :    : ||| ||| |||
Db      303 rkkaavlilgagrvcqpaaemlssfgrpsssqwyktlleddfecqtdvevivgslylkda 362

Qy      594 EETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTAS 653
            |:||:|| | |  |||| |  :|   :|||:||||||| | |  :|  |||||||:||||
Db      363 eqtvegipnvtgiqldvmdranlckyisqvdvvisllppschiivanacielkkhlvtas 422

Qy      654 YVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLP 713
            ||| ||| |:  |||||:||| |||||||| |:|:||||::|| ||||||:|||||||||
Db      423 yvdssmsmlndkakdagitilgemgldpgighmmamkminqahvrkgkiksftsycgglp 482

Qy      714 SPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFA 773
            || |||||||||||||||||:|:|:||| ||: |||:|:|| :||:|| |||| :|||||
Db      483 speaannplaykfswnpagairagrnpatykwggetvhidgddlydsatrlrlpdlpafa 542

Qy      774 LEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTS 833
            || |||||||:|||||||: |||||:| | ||||||||| |||:   |:   | || :
```

```
Db    543 leclpnrnsllygdlygit-eastifrgtlryegfseimgtlsrislfnneahsllmngq 601

Qy    834 RPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFL 893
          |||:| || |||  :     :| |      ::|:: ::|  | ||::  |::| ||| ||
Db    602 rptfkkflfellkvvgdnpdellig-----endimeqiliqghckdqrtametaktiifl 656

Qy    894 GLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATL 953
          || ::|:||  | | ||| | |||:|::|   |:||||||||||:|||| |  |||:|||
Db    657 glldqteipascksafdvacfrmeerlsytstekdmvllhheveieypdsqitekhratl 716

Qy    954 LEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSG 1013
          |||||  : ::|||||||||||||:|||||| ||:||:||:||::||:| |||:|:|: |
Db    717 lefgktldektttamaltvgipaavgalllltnkiqtrgvlrpiepevynpaldiieayg 776

Qy   1014 IKLVEKVE 1021
          |||:|| |
Db    777 ikliekte 784

RESULT   8
AAW87766
ID   AAW87766 standard; Protein; 482 AA.
XX
AC   AAW87766;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Arabidopsis saccharopine dehydrogenase.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed.
XX
OS   Arabidopsis thaliana.
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   N-PSDB; AAV99568.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 2; Page 201-202; 231pp; English.
XX
CC   This is the amino acid sequence of the saccharopine dehydrogenase
CC   (SDH) domain of the lysine ketoglutarate reductase (LKR)/SDH protein
CC   of Arabidopsis thaliana.  Nucleic acid fragments comprising
CC   sequences encoding all or part of plant LKR polypeptides are new.
CC   Also claimed are: (1) a chimeric gene comprising the fragment (or a
CC   subfragment) operably linked to a suitable seed specific regulator,
CC   where the chimeric gene reduces LKR activity in plant seeds
CC   transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
```

```
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence   482 AA;

  Query Match             28.0%;  Score 1480;  DB 20;  Length 482;
  Best Local Similarity   60.0%;  Pred. No. 6.7e-123;
  Matches  290;  Conservative   78;  Mismatches  101;  Indels   14;  Gaps    4;

Qy    546 ILILGAGRVCRPAAEFLASYPDICT-------YGVDDHDADQIHVIVASLYQKDAEETVD 598
          :|||||||||||||:||||   | :       :| |  :   :|||||||| |||:|||:
Db      7 vlilgagrvcrpaadflasvrtissqqwyktyfgadseektdvhvivaslylkdaketve 66

Qy    599 GIENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDES 658
          || :  | :|||:|  ||   ||||:||:|||||| || :|  |||||||:|||||||:
Db     67 gisdveavrldvsdsesllkyvsqvdvvlsllpaschavvaktcielkkhlvtasyvdde 126

Qy    659 MSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAA 718
           | | : || ||:||| ||||||||||:|:||||::|| :|||:|:|:|||||||||||||
Db    127 tsmlhekaksagitilgemgldpgidhmmamkmindahikkgkvksftsycgglpspaaa 186

Qy    719 NNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLP 778
          |||||||||||||||:|:|:||| ||  |: ||||| |||:|| | |:  |||||||  |
Db    187 nnplaykfswnpagairagqnpakyksngdiihvdgknlydsaarfrvpnlpafalecfp 246

Qy    779 NRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYK 838
          ||:||:||: |||  ||:||:| | |||||| || |||| ||||:  : :|    | |:
Db    247 nrdslvygehygieseattifrgtlryegfsmimatlsklgffdseanqvlstgkritfg 306

Qy    839 GFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHEE 898
            |  :||     : | : |   | ::::  |::|||   :|| | |  ||| ||| :||
Db    307 allsnilnk----dadneseplag-eeeiskriiklg--hsketaakaaktivflgfnee 359

Qy    899 TQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEFGK 958
           ::|  | | ||  |  ||:::||  |||||||||||||||: : :  ||| |||||||
Db    360 revpslcksvfdatcylmeeklaysgneqdmvllhhevevefleskriekhtatllefgd 419

Qy    959 VENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSGIKLVE 1018
          ::||::||||| |||||||||||:|:::|::|:||:|||: |:|:|||:||:: ||||:|
Db    420 ikngqtttamaktvgipaaigalvliedkiktrgvlrpleaevylpaldilqaygiklme 479

Qy   1019 KVE 1021
          | |
Db    480 kae 482

RESULT   9
AAE05075
ID   AAE05075 standard; Protein; 928 AA.
XX
AC   AAE05075;
XX
DT   12-SEP-2001  (first entry)
XX
DE   Drosophila melanogaster LKR/SDH protein.
XX
KW   Fruitfly; lysine-ketoglutarate reductase; LKR; pesticide; therapy;
KW   saccharopine dehydrogenase; SDH; invertebrate enzyme; drug screening;
KW   peptidyl-alpha-hydroxyglycine alpha-amidating lyase; PAL;
KW   genetic modification; lysine catabolism disorder.
XX
OS   Drosophila melanogaster.
XX
FH   Key             Location/Qualifiers
FT   Domain          23..474
FT                   /note= "Alanine dehydrogenase/pyridine nucleotide
FT                   transhydrogenase domain"
```

```
FT   Domain          485..498
FT                   /note= "Lactate/malate dehydrogenase domain"
XX
PN   WO200149856-A2.
XX
PD   12-JUL-2001.
XX
PF   05-JAN-2001; 2001WO-US00360.
XX
PR   06-JAN-2000; 2000US-0174973.
PR   29-FEB-2000; 2000US-0185844.
PR   22-MAR-2000; 2000US-0191189.
XX
PA   (GENO-) GENOPTERA LLC.
XX
PI   Stout TJ,  Keegan KP,  Ebens AJ;
XX
DR   WPI; 2001-441885/47.
DR   N-PSDB; AAD08977.
XX
PT   Novel invertebrate enzymes and nucleic acid encoding the enzyme useful
PT   as targets for pesticides and to identify compounds that have utility
PT   as therapeutics or pesticides  -
XX
PS   Claim 15; Page 49-51; 54pp; English.
XX
CC   The invention relates to invertebrate enzymes and nucleic acids,
CC   including lysine-ketoglutarate reductase/saccharopine dehydrogenase
CC   (LKR/SDH) and peptidyl-alpha-hydroxyglycine alpha-amidating lyase (PAL)
CC   from Drosophila melanogaster, designated dmLKR/SDH, dmPAL and dmPAL2.
CC   The enzymes are useful for detecting a candidate compound especially a
CC   putative pesticidal or pharmaceutical agent that interacts with an
CC   invertebrate enzyme or its fragment. The identified compound is useful
CC   for controlling a pest, by reducing viability of the pest. Insect or
CC   worm genetically modified to express the enzyme are useful for studying
CC   invertebrate enzyme activity and for identification and screening of
CC   pesticide targets directed to components of a pathway involving a
CC   subject protein. Nucleic acids encoding the invertebrate enzymes or their
CC   fragments are useful as biopesticides. The enzymes and their DNA are
CC   useful for genetically modifying metazoan invertebrate organisms, such
CC   as insects and worms or cultured cells, resulting in expression or
CC   mis-expression of the protein. The organisms or cultured cells are useful
CC   for screening candidate compounds for the ability to modulate expression
CC   or activity of a subject nucleic acid or protein, and thus are useful in
CC   identification of new drug targets, therapeutic agents, diagnostics and
CC   prognostics useful in treatment of disorders associated with abnormal
CC   enzyme levels and/or function, e.g. disorders associated with lysine
CC   catabolism. The present sequence is Drosophila melanogaster LKR/SDH
CC   (dmLKR/SDH) protein. LKR/SDH is a principal enzyme in the catabolism of
CC   lysine.
XX
SQ   Sequence   928 AA;

  Query Match             27.7%;  Score 1466.5;  DB 22;  Length 928;
  Best Local Similarity   33.4%;  Pred. No. 3.2e-121;
  Matches  343;  Conservative  183;  Mismatches  318;  Indels  183;  Gaps   22;

Qy      17 RIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQMILSDRAYAFFSHTHKA 76
           ::||||| || :    |  ||  | ||:|:  :| |:||  :  ::  : | ||||| ||
Db      54 kvivqpsnrraypmqaymqagahiqedisdasvifgvkqvpidalipgktycffshtika 113

Qy      77 QKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRYLSLGY 136
           |: |||||| |||::: | ||| |: : | | :||||:|| ||::: |||:| | |:||:
Db     114 qesnmplldailekkirlidyeriidergarqvafgkyagvagmvnilhgiglrllalgh 173

Qy     137 STPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVSQGAQEIF 196
            |||: :| :| | : : |: |:     ||:   :|  | |: ||||| |||||||||:|
Db     174 htpfmhigpahnyrnssmarqairdcgyeislgmmpksigpltfvftgsgnvsqgaqevf 233

Qy     197 KLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDIVSHKDPTRQFDKGDY 256
             ||  :|  | | :: :  | :|          |||| |:  | :  ::    ||  :|
Db     234 selpieyvppemlrkvaehgnqnk----------lygcevsrsdhlerre-gggfdakey 282

Qy     257 YAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLM----------ETGCP 306
              || |   |  :|||||||||| :||    | |:::   : |:            | |
Db     283 defperyistfstkiapyasvivngiywavgspklisipdaknllrpantpwlptskgsp 342

Qy     307 -----LVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDIL 361
                :: :|||: | ||||||:|: |:|: ||  ||  :|      :| ||:  ::| :
Db     343 alphrmlaicdisadpggsiefmnecttidtpfclydadrnkdtksfkgpgvlvcsidnm 402

Qy     362 PTEFSKEASQHFGNILSRLVASLASVKQPAELP------SY-LRRACIAHAGRLTPLYEY 414
           ||:  :|::  || :|:  |  :  :|  |: |      || :: | ||  |:||  ::|
Db     403 ptqlprestdlfgellaphvhdi--iksdakkplaeenfsypiqsaiiasngqltegfqy 460

Qy     415 IPRMRNTMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCE 474
           |  :|
Db     461 iqelr------------------------------------------------------- 465

Qy     475 VGQSTDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYE 534
             :|    |  ::| |: ::
Db     466 --esqshrsrhkme-gsses---------------------------------------- 482

Qy     535 TDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAE 594
                ||   |:|:|||| |  |  |:|             |    : : | |  :::|:
Db     483 -----dk---kvlvlgagmvsaplvewl-------------hrekdvsitvcsqvkeead 521

Qy     595 ETVDGIENTTATQLDVAD-IGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTAS 653
                     :  ||| :  | | :|  : :||:|||| | |  :|  |:    ||||||
Db     522 rlaqqyagvdsvyldvnestghlqelcgradvvvsllpyslhgmvarycvaegthmvtas 581

Qy     654 YVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLP 713
           |::: :| | : ||  ||||: |:|||||||||:::: | |   :   :::| |||||||
Db     582 ylndeisglheeakakgvtimnevgldpgidhllalecihevqdkgavvesfvsycgglp 641

Qy     714 SPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFA 773
           :|  :|| | |||||:| | | :  : | |   |: : : |     |: |  |  || ||
Db     642 apehsnnalrykfswsprgvllntlsaakylsqgqiveisgggelmsspr-sldflpgfa 700

Qy     774 LEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTS 833
           ||  |||:|  ||:|||: ::  |: | | ||:|||| :  :   |  |   | ||  :
Db     701 legfpnrdstkygnlyglgrdvhtllrgtirykgfsesikpmqllglidpephallhpsg 760

Qy     834 RP-TYKGFLDELLN-NISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIK 891
              |::  :  |:  : |||           : ::|  :| :          :  |  |:
Db     761 pdvtwrqlvihlmgmsdsti-----------fyenlkqklte---------rigdvdgie 800

Qy     892 FLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQA 951
            ||| ::| : |  ::| | :   : :|:|:  :|:|:|:| ||| : :|||:  |:
Db     801 slgllddtpvvk-lntpldtlshylskrlaferderdlvvlrhevgirwpdgr-reergi 858

Qy     952 TLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILES 1011
             : :|: : | |  |||:||| |||| | ::|  ::| :||: |  |:|| | |: | |
Db     859 nfvvygqpq-ghs--amamtvgkpaaiaakmildgeiqergvllpftpdiyrpmlqrlrs 915

Qy    1012 SGIKLVE 1018
            |:   |
Db     916 egltate 922


RESULT  10
ABB63015
ID   ABB63015 standard; Protein; 972 AA.
XX
AC   ABB63015;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 15837.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
```

```
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07118.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 15837; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   972 AA;


  Query Match             27.7%;  Score 1466.5;  DB 22;  Length 972;
  Best Local Similarity   33.4%;  Pred. No. 3.4e-121;
  Matches  343;  Conservative  183;  Mismatches  318;  Indels  183;  Gaps   22;

Qy      17 RIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQMILSDRAYAFFSHTHKA 76
           ::||||| || :    |  ||  | ||:|:  :| |:||  :  ::  : | ||||| ||
Db      98 kvivqpsnrraypmqaymqagahiqedisdasvifgvkqvpidalipgktycffshtika 157

Qy      77 QKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRYLSLGY 136
           |: |||||| |||::: | ||| |: : | | :||||:|| ||::: |||:| | |:||:
Db     158 qesnmplldailekkirlidyeriidergarqvafgkyagvagmvnilhgiglrllalgh 217

Qy     137 STPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVSQGAQEIF 196
            |||: :| :| | : : |: |:     ||:   :|  | |: ||||| ||||||||:|
Db     218 htpfmhigpahnyrnssmarqairdcgyeislgmmpksigpltfvftgsgnvsqgaqevf 277

Qy     197 KLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDIVSHKDPTRQFDKGDY 256
             ||  :|  | | :: :  | :|          |||| |:  | :  ::    ||  :|
Db     278 selpieyvppemlrkvaehgnqnk----------lygcevsrsdhlerre-gggfdakey 326

Qy     257 YAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLM----------ETGCP 306
              || |   |  :||||||||||| :||    | |:::   : |:            | |
Db     327 defperyistfstkiapyasvivngiywavgspklisipdaknllrpantpwlptskgsp 386

Qy     307 -----LVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDIL 361
                :: :|||: | ||||||:|: |:|: ||  ||  :|      :| ||:  ::| :
Db     387 alphrmlaicdisadpggsiefmnecttidtpfclydadrnkdtksfkgpgvlvcsidnm 446

Qy     362 PTEFSKEASQHFGNILSRLVASLASVKQPAELP------SY-LRRACIAHAGRLTPLYEY 414
           ||:  :|::  || :|:  |  :  :|  |: |      || :: | ||  |:||  ::|
Db     447 ptqlprestdlfgellaphvhdi--iksdakkplaeenfsypiqsaiiasngqltegfqy 504

Qy     415 IPRMRNTMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCE 474
           |  :|
Db     505 iqelr------------------------------------------------------- 509
```

```
Qy     475 VGQSTDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYE 534
             :|    |  ::| |: ::
Db     510 --esqshrsrhkme-gsses---------------------------------------- 526

Qy     535 TDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAE 594
                ||   |:|:|||| |  |  |:|               |    : : | |  :::|:
Db     527 -----dk---kvlvlgagmvsaplvewl-------------hrekdvsitvcsqvkeead 565

Qy     595 ETVDGIENTTATQLDVAD-IGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTAS 653
                     :  ||| :  | | :|  : :||:|||| | |  :|  |:    ||||||
Db     566 rlaqqyagvdsvyldvnestghlqelcgradvvvsllpyslhgmvarycvaegthmvtas 625

Qy     654 YVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLP 713
           |::: :| | : ||  ||||: |:|||||||||:::: | |   :   :::| |||||||
Db     626 ylndeisglheeakakgvtimnevgldpgidhllalecihevqdkgavvesfvsycgglp 685

Qy     714 SPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFA 773
           :|  :|| | |||||:| | | :  : | |   |: : : |     |: |  |  || ||
Db     686 apehsnnalrykfswsprgvllntlsaakylsqggiveisgggelmsspr-sldflpgfa 744

Qy     774 LEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTS 833
           ||  |||:|  ||:|||: ::  |: | | ||:|||| :  :   |  |   | ||  :
Db     745 legfpnrdstkygnlyglgrdvhtllrgtirykgfsesikpmqllglidpephallhpsg 804

Qy     834 RP-TYKGFLDELLN-NISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIK 891
              |::  :  |:  : |||            : ::|  :| :          :  |  |:
Db     805 pdvtwrqlvihlmgmsdsti-----------fyenlkqklte---------rigdvdgie 844

Qy     892 FLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQA 951
            ||| ::| : |  ::| | :   : :|:|::  :|:|:|:|| ||| : :|||:  |:
Db     845 slglldd tpvvk-lntpldtlshylskrlaferderdlvvlrhevgirwpdgr-reergi 902

Qy     952 TLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILES 1011
             : :|: : | |  |||:||| |||| | ::|  ::| :||: |  |:|| | |: | |
Db     903 nfvvygqpq-ghs--amamtvgkpaaiaakmildgeiqergvllpftpdiyrpmlqrlrs 959

Qy    1012 SGIKLVE 1018
            |:   |
Db     960 egltate 966


RESULT  11
AAW87763
ID   AAW87763 standard; Protein; 239 AA.
XX
AC   AAW87763;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Rice lysine ketoglutarate reductase partial sequence.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; rice.
XX
OS   Oryza sativa.
XX
FH   Key               Location/Qualifiers
FT   Misc-difference   72
FT                     /note= "encoded by NAT"
FT   Misc-difference   155
FT                     /note= "encoded by MAG"
FT   Misc-difference   226
FT                     /note= "encoded by ANG"
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
```

```
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   N-PSDB; AAV99565.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 2; Page 196-197; 231pp; English.
XX
CC   This is the amino acid sequence of a rice lysine ketoglutarate
CC   reductase (LKR) or saccharopine dehydrogenase (SDH) partial
CC   polyeptide.  It was deduced from cDNA (see AAV99565) obtained from
CC   an EST database search.  Isolated nucleic acid fragments
CC   comprising sequences encoding all or part of plant LKR enzymes
CC   are new.  Also claimed are: (1) a chimeric gene comprising the
CC   fragment (or a subfragment) operably linked to a seed specific
CC   regulator, where the chimeric gene reduces LKR activity in plant
CC   seeds transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence   239 AA;

  Query Match             18.8%;  Score 996;  DB 20;  Length 239;
  Best Local Similarity   83.2%;  Pred. No. 3.1e-80;
  Matches  188;  Conservative   21;  Mismatches   17;  Indels    0;  Gaps    0;

Qy    616 LSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILC 675
          ||:||||||||:|||||||||||| ||||:|||:||||||||||| | |:|: |||||||
Db      4 lsnlvsqvevvvsllpasfhaaiarvciemkkhlvtasyvdesmskleqsaegagvtilc 63

Qy    676 EMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALR 735
          |||||||| |:||||||||||:||||||:|||:||||||||:|||||||||||:||||:|
Db     64 emgldpgixhmmsmkmideahsrkgkiksftsfcgglpspasannplaykfswspagair 123

Qy    736 SGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEA 795
          :|:||||||| || |||||  |||||||||| ||||||||||||||||:|||||||||||
Db    124 agrnpavykfhgeiihvdgdklyesakrlrlxelpafalehlpnrnslmygdlygiskea 183

Qy    796 STIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGFL 841
          ||:|||| |||||:||| | :| ||||||:||||| |:||||  ||
Db    184 stvyratlryegfneimatfakigffdaashpllqqttrptyxdfl 229


RESULT  12
AAW87765
ID   AAW87765 standard; Protein; 83 AA.
XX
AC   AAW87765;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Wheat lysine ketoglutarate reductase partial sequence.
XX
```

```
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed; wheat
XX
OS   Triticum aesitvum.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 58
FT                   /note= "encoded by NTG"
FT   Misc-difference 62
FT                   /note= "encoded by ATN"
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   N-PSDB; AAV99567.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 2; Page 199; 231pp; English.
XX
CC   This is the amino acid sequence of a wheat lysine ketoglutarate
CC   reductase (LKR) or saccharopine dehydrogenase (SDH) partial
CC   polyeptide.  It was deduced from cDNA (see AAV99567) obtained from
CC   an EST database search.  Isolated nucleic acid fragments
CC   comprising sequences encoding all or part of plant LKR enzymes
CC   are new.  Also claimed are: (1) a chimeric gene comprising the
CC   fragment (or a subfragment) operably linked to a seed specific
CC   regulator, where the chimeric gene reduces LKR activity in plant
CC   seeds transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence   83 AA;

  Query Match              6.8%;  Score 360;  DB 20;  Length 83;
  Best Local Similarity   83.1%;  Pred. No. 2.9e-24;
  Matches   69;  Conservative    7;  Mismatches    7;  Indels    0;  Gaps    0;

Qy    940 YPDGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQP 999
          |||| | |||||||||||| |||| |||||||||:||||||||||:|||| |||||| :|
Db      1 ypdgdptekhqatllefgktengrpttamaltvgvpaaigallllqnkvqrkgvirpxep 60

Qy   1000 EIYVPALEILESSGIKLVEKVET 1022
          | |:|||||||:|||||:|:|||
Db     61 exyipaleileasgiklierveT 83
```

```
RESULT  13
AAW60533
ID   AAW60533 standard; Protein; 74 AA.
XX
AC   AAW60533;
XX
DT   25-AUG-1998  (first entry)
XX
DE   Protein homologous to fungal saccharopine dehydrogenases.
XX
KW   Dihydrodipicolinic acid synthase; DHDPS; chimeric gene;
KW   lysine inhibition; plant chloroplast transit sequence;
KW   plant seed-specific regulatory sequence; transgenic plant;
KW   increased lysine level; corn; Zea mays; soybean; Glycine max.
XX
OS   Arabidopsis thaliana.
XX
PN   US5773691-A.
XX
PD   30-JUN-1998.
XX
PF   07-JUN-1995;   95US-0474633.
XX
PR   19-MAR-1992;   92US-0855414.
PR   18-MAR-1993;   93WO-US02480.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Falco SC,  Keeler SJ,  Rice JA;
XX
DR   WPI; 1998-387117/33.
DR   N-PSDB; AAV35857.
XX
PT   Chimeric genes encoding lysine production enzymes - useful for
PT   increasing transgenic seed lysine content without being inhibited by
PT   high levels of the amino acid
XX
PS   Example 20; Columns 155-156; 106pp; English.
XX
CC   The present sequence represents a protein homologous to fungal
CC   saccharopine dehydrogenases. The specification describes a
CC   Corynebacterium dapA gene, which encodes a dihydrodipicolinic acid
CC   synthase (DHDPS) enzyme, which was used to create chimeric genes of
CC   the invention. The chimeric genes contain a nucleic acid fragment
CC   encoding a DHDPS enzyme which is insensitive to inhibition by lysine
CC   operably linked to a plant chloroplast transit sequence and to a plant
CC   seed-specific regulatory sequence. The chimeric genes are useful for
CC   producing plants containing increased levels of lysine, especially in
CC   corn (Zea mays) and soybean (Glycine max).
XX
SQ   Sequence   74 AA;

  Query Match             5.1%;  Score 270;  DB 19;  Length 74;
  Best Local Similarity   67.6%;  Pred. No. 2.5e-16;
  Matches   50;  Conservative   18;  Mismatches    6;  Indels    0;  Gaps    0;

Qy     948 KHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALE 1007
           || ||||||| ::||::||||| |||||||||||||:::|::|:||:|||: |:|:|||:
Db       1 khtatllefgdikngqtttamaktvgipaaigallliedkiktrgvlrpleaevylpald 60

Qy    1008 ILESSGIKLVEKVE 1021
           ||:: ||||:|| |
Db      61 ilqaygiklmekae 74


RESULT  14
AAW87758
ID   AAW87758 standard; Protein; 74 AA.
XX
AC   AAW87758;
```

```
XX
DT   29-MAR-1999  (first entry)
XX
DE   Arabidopsis lysine ketoglutarate reductase.
XX
KW   Lysine ketoglutarate reductase; saccharopine dehydrogenase;
KW   transgenic plant; seed.
XX
OS   Arabidopsis thaliana.
XX
PN   WO9842831-A2.
XX
PD   01-OCT-1998.
XX
PF   27-MAR-1998;   98WO-US06051.
XX
PR   27-MAR-1997;   97US-0824627.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Epelbaum SU,  Falco SC,  McDevitt RE;
XX
DR   WPI; 1999-045139/04.
DR   N-PSDB; AAV99554.
XX
PT   Nucleic acids and chimeric genes for increasing seed lysine content
PT   - comprise sequence encoding all or part of lysine ketoglutarate
PT   reductase, useful to improve nutritional quality of seeds from
PT   transformed plants
XX
PS   Claim 2; Page 166-167; 231pp; English.
XX
CC   This is the amino acid sequence of an Arabidopsis thaliana
CC   lysine ketoglutarate reductase (LKR) polypeptide, identified on the
CC   basis of homology to a fungal glutamate-forming saccharopine
CC   dehydrogenase.  Isolated nucleic acid fragments comprising
CC   sequences encoding all or part of plant LKR polypeptides are new.
CC   Also claimed are: (1) a chimeric gene comprising the fragment (or a
CC   subfragment) operably linked to a suitable seed specific regulator,
CC   where the chimeric gene reduces LKR activity in plant seeds
CC   transformed with it; (2) plant cells and seeds in which LKR
CC   activity is reduced due to a mutation in the gene encoding LKR or
CC   transformation with the chimeric gene; (3) a nucleic acid fragment
CC   comprising: (i) chimeric gene above, and (ii) a second chimeric gene,
CC   in which a nucleic acid fragment encoding dihydrodipicolinic acid
CC   synthase (DHDPS) substantially insensitive to lysine inhibition is
CC   operably linked to a plant chloroplast transit sequence and to a
CC   plant seed-specific regulatory sequence; (4) plants comprising in
CC   genome (i) and (ii), especially as fragment of (2); and (5) seeds
CC   from (3).  The chimeric genes can be used to produce plant cells
CC   and seeds with reduced LKR activity, especially in Arabidopsis,
CC   corn, soybean, rapeseed, wheat and rice (claimed).  LKR is
CC   important in controlling free lysine accumulation in plant seeds.
CC   LKR activity reduction may be achieved by cloning the claimed
CC   fragment, preparing a chimeric gene for cosuppression of LKR,
CC   expression of antisense RNA for LKR, and transforming plants with
CC   the chimeric gene.
XX
SQ   Sequence   74 AA;

  Query Match             5.1%;  Score 270;  DB 20;  Length 74;
  Best Local Similarity   67.6%;  Pred. No. 2.5e-16;
  Matches   50;  Conservative   18;  Mismatches    6;  Indels    0;  Gaps    0;

Qy     948 KHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALE 1007
           || ||||||| ::||::||||| |||||||||||||:::|::|:||:|||: |:|:|||:
Db       1 khtatllefgdikngqtttamaktvgipaaigallliedkiktrgvlrpleaevylpald 60

Qy    1008 ILESSGIKLVEKVE 1021
           ||:: ||||:|| |
Db      61 ilqaygiklmekae 74
```

```
RESULT  15
AAW60532
ID   AAW60532 standard; Protein; 123 AA.
XX
AC   AAW60532;
XX
DT   25-AUG-1998  (first entry)
XX
DE   Protein homologous to fungal saccharopine dehydrogenases.
XX
KW   Dihydrodipicolinic acid synthase; DHDPS; chimeric gene;
KW   lysine inhibition; plant chloroplast transit sequence;
KW   plant seed-specific regulatory sequence; transgenic plant;
KW   increased lysine level; corn; Zea mays; soybean; Glycine max.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 32
FT                   /note= "not specified"
XX
PN   US5773691-A.
XX
PD   30-JUN-1998.
XX
PF   07-JUN-1995;   95US-0474633.
XX
PR   19-MAR-1992;   92US-0855414.
PR   18-MAR-1993;   93WO-US02480.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Falco SC,  Keeler SJ,  Rice JA;
XX
DR   WPI; 1998-387117/33.
DR   N-PSDB; AAV35856.
XX
PT   Chimeric genes encoding lysine production enzymes - useful for
PT   increasing transgenic seed lysine content without being inhibited by
PT   high levels of the amino acid
XX
PS   Example 20; Columns 155-156; 106pp; English.
XX
CC   The present sequence represents a protein homologous to fungal
CC   saccharopine dehydrogenases. The specification describes a
CC   Corynebacterium dapA gene, which encodes a dihydrodipicolinic acid
CC   synthase (DHDPS) enzyme, which was used to create chimeric genes of
CC   the invention. The chimeric genes contain a nucleic acid fragment
CC   encoding a DHDPS enzyme which is insensitive to inhibition by lysine
CC   operably linked to a plant chloroplast transit sequence and to a plant
CC   seed-specific regulatory sequence. The chimeric genes are useful for
CC   producing plants containing increased levels of lysine, especially in
CC   corn (Zea mays) and soybean (Glycine max).
XX
SQ   Sequence   123 AA;


  Query Match             5.0%;  Score 263.5;  DB 19;  Length 123;
  Best Local Similarity   51.6%;  Pred. No. 2.2e-15;
  Matches   64;  Conservative   17;  Mismatches   26;  Indels   17;  Gaps    4;

Qy    526 KIGKVNEYETDVTIDKGGPK------ILILGAGRVCRPAAEFLASYPDICT-------YG 572
          ||||| : | ::   |  |:      :|||||||| ||||:||||   | :       :|
Db      1 kigkvqq-enei---kekpemtkksgvlilgagrvxrpaadflasvrtissqqwyktyfg 56

Qy    573 VDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLPA 632
           |  :   :|||||||| |||:|||:|| :  | :|||:|  ||   ||||:||:|||||
Db     57 adseektdvhvivaslylkdaketvegisdveavrldvsdsesllkyvsqvdvvlsllpa 116

Qy    633 SFHA 636
          | ||
Db    117 scha 120
```

```
Search completed: May 31, 2002, 15:47:09
Job time: 151 sec
```

```
                        GenCore version 4.5
                Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         May 31, 2002, 15:44:58 ; Search time 26.05 Seconds
                                           (without alignments)
                                           3769.800 Million cell updates/sec

Title:          US-09-049-304A-122
Perfect score:  5286
Sequence:       1 CARLLLGGGKNGPRVNRIIV..........VPALEILESSGIKLVEKVET 1022

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters:      283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_71:*
               1:  pir1:*
               2:  pir2:*
               3:  pir3:*
               4:  pir4:*

       Pred. No. is the number of results predicted by chance to have a
       score greater than or equal to the score of the result being printed,
       and is derived by analysis of the total score distribution.

                                SUMMARIES
                     %
Result             Query
  No.     Score    Match  Length  DB  ID                      Description
-----------------------------------------------------------------------------
    1      4750     89.9    1056   2  T02930                  lysine-ketoglutara
    2      3411     64.5    1064   2  T05195                  saccharopine dehyd
    3    1429.5     27.0     934   2  T15063                  hypothetical prote
    4       716     13.5     450   2  T40337                  probable saccharop
    5     672.5     12.7     446   2  S41937                  saccharopine dehyd
    6       608     11.5     177   2  T07843                  probable lysine-ke
    7     216.5      4.1     352   2  A71176                  hypothetical prote
    8       172      3.3     369   2  A36467                  saccharopine dehyd
    9       157      3.0     368   2  T50174                  probable saccharop
   10     154.5      2.9     373   2  S48496                  saccharopine dehyd
   11     148.5      2.8     366   2  F97474                  hypothetical prote
   12     148.5      2.8     366   2  AB2693                  dehydrogenase Atu0
   13     145.5      2.8    1825   2  S13507                  microtubule-associ
   14     141.5      2.7    1830   2  A37981                  microtubule-associ
   15     139.5      2.6     605   2  I39837                  dnaK-type molecula
   16     136.5      2.6     414   2  A82177                  conserved hypothet
   17     136.5      2.6     703   2  AC2430                  hypothetical prote
   18       135      2.6     405   2  B75420                  conserved hypothet
   19     133.5      2.5    1300   2  T00317                  probable serine pr
   20       132      2.5     398   2  S74347                  hypothetical prote
   21     130.5      2.5     419   2  D97974                  conserved hypothet
   22       130      2.5     407   2  E69409                  conserved hypothet
   23     127.5      2.4     411   2  G69215                  conserved hypothet
   24     127.5      2.4     419   2  B95106                  conserved hypothet
   25       127      2.4     705   2  S76729                  hypothetical prote
   26       127      2.4    1068   2  T48756                  mitochondrial nico
   27     125.5      2.4    1451   2  A36468                  SPT6 protein - yea
   28       125      2.4    2261   2  T20978                  hypothetical prote
   29       124      2.3    5255   2  T31677                  bacitracin synthet
   30       123      2.3     408   2  C69448                  conserved hypothet
   31       122      2.3    1311   2  T08986                  hypothetical prote
   32       122      2.3    4717   2  T41581                  hypothetical coile
   33       121      2.3     760   2  T34414                  hypothetical prote
   34       121      2.3    2672   2  A48126                  translation activa
   35       120      2.3     493   2  A26941                  cholesteryl ester
   36       120      2.3    2241   2  T20971                  hypothetical prote
   37       119      2.3     401   2  C81435                  hypothetical prote
   38       119      2.3    1828   2  A40115                  microtubule-associ
   39     118.5      2.2     360   2  S74638                  alanine dehydrogen
   40     118.5      2.2     384   2  AG1948                  hypothetical prote
   41       118      2.2     428   2  A45732                  phosphopyruvate hy
   42     117.5      2.2     946   2  A71805                  probable ATP-depen
   43     116.5      2.2     399   2  H71810                  hypothetical prote
   44     116.5      2.2    1450   2  S78060                  probable DNA-direc
   45     116.5      2.2    2802   2  F97686                  cyclic beta-(1-2)

                                ALIGNMENTS

RESULT   1
T02930
lysine-ketoglutarate reductase / saccharopine dehydrogenase (NADP+, L-glutamate-formi
N;Contains: lysine-ketoglutarate reductase; saccharopine dehydrogenase (NADP+, L-glut
C;Species: Zea mays (maize)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T02930
R;Cord-Neto, G.; Kemper, E.L.; Arruda, P.
submitted to the EMBL Data Library, May 1998
A;Description: Lysine-ketoglutarate reductase/saccharopine dehydrogenase bifunctional
A;Reference number: Z14777
A;Accession: T02930
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1056 <COR>
A;Cross-references: EMBL:AF003551; NID:g3157909; PID:g3157910
A;Experimental source: cultivar AGROCERES F-352 commercial hybrid
C;Keywords: oxidoreductase

  Query Match             89.9%;  Score 4750;  DB 2;  Length 1056;
  Best Local Similarity   92.0%;  Pred. No. 0;
  Matches  941;  Conservative   18;  Mismatches   58;  Indels    6;  Gaps    5;

Qy      1 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     39 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 98

Qy     61 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     99 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGTAGL 158

Qy    121 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 218

Qy    181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRD 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRD 278

Qy    241 IVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQL 300
          :||||  : |    ||||||:  || |||||||||||||:||||||||||||||||||||
Db    279 MVSHKG-SHQTLTSDYYAHPD-TTPCFHERIAPYASVIVHCMYWEKRFPPLLNMDQLQQL 336

Qy    301 METGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDI 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    337 METGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDI 396

Qy    361 LPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMRN 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    397 LPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMRN 456
```

```
Qy    421 TMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQSTD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    457 TMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQSTD 516

Qy    481 DMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGKVNEYETDVTID 540
          ||||||||||||||||||||||||||||||| |:  :    |||||||||        |
Db    517 DMSYSELEVGADDTATLDKIIDSLTSLANEH-GEITSRARNELALKIGKVMNMNDSNVIK 575

Qy    541 KGGPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGI 600
          :|  : || ||       ||||||||||||||||||||||||||||||||||||||||||
Db    576 EG--QDLIFGAEESVGTTAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGI 633

Qy    601 ENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMS 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    634 ENTTATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMS 693

Qy    661 NLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANN 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    694 NLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANN 753

Qy    721 PLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNR 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    754 PLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNR 813

Qy    781 NSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGF 840
          |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    814 NSLIYGDLYGISKEASTIYRATLRYEGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGF 873

Qy    841 LDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHEETQ 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    874 LDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHEETQ 933

Qy    901 IPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYP-DGQPAEKHQATLLEFGKV 959
          || |||||||||||   :     :  ||||||||||||:    | | :  :      |:|
Db    934 IPTGCSSPFDVICQEWNRGWPMAYMSQDMVLLHHEVEVDITRTGNPPKSTKRRNGVSGRV 993

Qy    960 ENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSGIKLVEK 1019
          |||||||||||||||||||||||||:|||||||:||||||||||||||||||||||||||
Db    994 ENGRSTTAMALTVGIPAAIGALLLLQNKVQTKGMIRPLQPEIYVPALEILESSGIKLVEK 1053

Qy   1020 VET 1022
          |::
Db   1054 VKS 1056


RESULT   2
T05195
saccharopine dehydrogenase (NADP+, L-lysine-forming) (EC 1.5.1.8) - Arabidopsis thaliana
N;Alternate names: lysine-ketoglutarate reductase; protein F4I10.80
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 23-Jul-1999
C;Accession: T05195
R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Hoheisel, J.; Mew
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15402
A;Accession: T05195
A;Molecule type: DNA
A;Residues: 1-1064 <BEV>
A;Cross-references: EMBL:AL035525
A;Experimental source: cultivar Columbia; BAC clone F4I10
C;Genetics:
A;Map position: 4
A;Introns: 93/3; 122/3; 169/2; 224/1; 251/3; 290/3; 316/1; 373/2; 405/3; 461/2; 481/3; 5
A;Note: F4I10.80
C;Keywords: oxidoreductase


  Query Match             64.5%;  Score 3411;  DB 2;  Length 1064;
  Best Local Similarity   63.6%;  Pred. No. 1e-221;
  Matches  668;  Conservative  151;  Mismatches  179;  Indels   52;  Gaps   13;

Qy      1 CARLLLGGGKNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQM 60
          ||| || |||:   ::||:|||| :|||||| ||| |||||:|||:||||:|||||:|:|
Db     38 CAR-LLHGGKDRTGISRIVVQPSAKRIHHDALYEDVGCEISDDLSDCGLILGIKQPELEM 96

Qy     61 ILSDRAYAFFSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGL 120
          || :|||||||||||||||||||||||| |||:| |||||||| ||| |||||:||||||
Db     97 ILPERAYAFFSHTHKAQKENMPLLDKILSERVTLCDYELIVGDHGKRLLAFGKYAGRAGL 156

Qy    121 IDFLHGLGQRYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
          :||||||||| | ||||||||||| |:|| |||||||||| | ||||: ||| ||||:||
Db    157 VDFLHGLGQRKLILGYSTPFLSLGASYMYSSLAAAKAAVISVGEEIASQGLPLGICPLVF 216

Qy    181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIF-QARNLSKQSQSTKRVFQLYGCVVTSR 239
          |||| |||| ||||||||||||||:  ||||:| : : :|:   |||||:|:|||::||:
Db    217 VFTGTGNVSLGAQEIFKLLPHTFVEPSKLPELFVKDKGISQNGISTKRVYQVYGCIITSQ 276

Qy    240 DIVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQ 299
          |:| ||||:: ||| ||||||||| |||||:|:|| ||:||||||||||| ||:  |||
Db    277 DMVEHKDPSKSFDKADYYAHPEHYNPVFHEKISPYTSVLVNCMYWEKRFPCLLSTKQLQD 336

Qy    300 LMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVD 359
          | : | ||||:|||||||||||||:|::| |: ||||::|| |||:|||:| ||:|:|||
Db    337 LTKKGLPLVGICDITCDIGGSIEFVNRATLIDSPFFRFNPSNNSYYDDMDGDGVLCMAVD 396

Qy    360 ILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMR 419
          ||||||:||||||||:|||  | ||||: : ::||::|:||||:: | || |||||||||
Db    397 ILPTEFAKEASQHFGDILSGFVGSLASMTEISDLPAHLKRACISYRGELTSLYEYIPRMR 456

Qy    420 NTMIDLAPAKTNP------------LPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGS 467
                   |:||               : :: |||||||||||||||||||:|| ||||
Db    457 ---------KSNPEEAQDNIIANGVSSQRTFNILVSLSGHLFDKFLINEALDMIEAAGGS 507

Qy    468 FHLVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLAN---EHGGDHDAGQEIELA 524
          ||| :||:||| |  |||||||||||   ||:|||||| |||   ::   |    :|  :
Db    508 FHLAKCELGQSADAESYSELEVGADDKRVLDQIIDSLTRLANPNEDYISPHREANKI--S 565

Qy    525 LKIGKVNEYETDVTIDKGGPK------ILILGAGRVCRPAAEFLASYPDICT-------Y 571
          |||||| : | ::   |  |:      :|||||||||||||:||||   | :       :
Db    566 LKIGKVQQ-ENEI---KEKPEMTKKSGVLILGAGRVCRPAADFLASVRTISSQQWYKTYF 621

Qy    572 GVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLLP 631
          | |  :   :|||||||| |||:|||:|| :  | :|||:|  ||   ||||:||:||||
Db    622 GADSEEKTDVHVIVASLYLKDAKETVEGISDVEAVRLDVSDSESLLKYVSQVDVVLSLLP 681

Qy    632 ASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKM 691
          || || :|  |||||||:|||||||:  | | : || ||:||| ||||||||||:|:|||
Db    682 ASCHAVVAKTCIELKKHLVTASYVDDETSMLHEKAKSAGITILGEMGLDPGIDHMMAMKM 741

Qy    692 IDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIH 751
          |::|| :|||:|:|||||||||||||||||||||||||||||:|:|:||| ||  |: ||
Db    742 INDAHIKKGKVKSFTSYCGGLPSPAAANNPLAYKFSWNPAGAIRAGQNPAKYKSNGDIIH 801

Qy    752 VDGHNLYESAKRLRLRELPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEI 811
          ||| |||:|| | |:  |||||||  |||:||:||: |||  ||:||:| | |||||| |
Db    802 VDGKNLYDSAARFRVPNLPAFALECFPNRDSLVYGEHYGIESEATTIFRGTLRYEGFSMI 861

Qy    812 MVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNISTINTDLDIEASGGYDDDLIARL 871
          | |||| ||||:  : :|    | |:   |  :||     : | : |   | ::::  |:
Db    862 MATLSKLGFFDSEANQVLSTGKRITFGALLSNILNK----DADNESEPLAG-EEEISKRI 916

Qy    872 LKLGCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVL 931
          :|||   :|| | |  ||| ||| :|| ::|  | | || |  ||:::||  |||||||
Db    917 IKLG--HSKETAAKAAKTIVFLGFNEEREVPSLCKSVFDATCYLMEEKLAYSGNEQDMVL 974

Qy    932 LHHEVEVEYPDGQPAEKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTK 991
          ||||||||: : :  ||| ||||||| ::||::||||| |||||||||||||:::|::|:
Db    975 LHHEVEVEFLESKRIEKHTATLLEFGDIKNGQTTTAMAKTVGIPAAIGALLLIEDKIKTR 1034

Qy    992 GVIRPLQPEIYVPALEILESSGIKLVEKVE 1021
          ||:|||: |:|:|||:||:: ||||:|| |
Db   1035 GVLRPLEAEVYLPALDILQAYGIKLMEKAE 1064


RESULT   3
T15063
```

```
hypothetical protein R02D3.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15063
R;Wu, X.; Antoniou, B.
submitted to the EMBL Data Library, December 1997
A;Description: The sequence of C. elegans cosmid R02D3.
A;Reference number: Z18284
A;Accession: T15063
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-934 <WUX>
A;Cross-references: EMBL:AF038615; NID:g2736322; PID:g2736323; PIDN:AAB94142.1; GSPDB:GN
A;Experimental source: strain Bristol N2; clone R02D3
C;Genetics:
A;Gene: CESP:R02D3.1
A;Map position: 4
A;Introns: 35/2; 135/3; 262/1; 304/3; 340/3; 476/2; 555/2; 599/2; 745/3; 857/1

  Query Match             27.0%;  Score 1429.5;  DB 2;  Length 934;
  Best Local Similarity   32.5%;  Pred. No. 4.6e-88;
  Matches  334;  Conservative  190;  Mismatches  321;  Indels  183;  Gaps   24;

Qy     10 KNGPRVNRIIVQPSTRRIHHDAQYEDAGCEISEDLSECGLIIGIKQPKLQMILSDRAYAF 69
          |:|  || :::||| || :    |  ||  : |||||  :|: :|   :  :: :: |||
Db     58 KSG--VN-VLIQPSNRRAYPIQDYISAGAIVREDLSEAHIIMSVKSVPIDQLIPNKTYAF 114

Qy     70 FSHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQ 129
          |||| |||::|| :|| ||:  : | ||| |  | ||| : |||:|| || || |||:|
Db    115 FSHTIKAQQDNMEMLDTILQRNIRLLDYEKICDDKGKRLVMFGKWAGNAGFIDILHGIGL 174

Qy    130 RYLSLGYSTPFLSLGQSHMYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVFTGVGNVS 189
          | |:||::|||| :| :| |     |  |:  :  |||   :|  : |::||||| ||||
Db    175 RLLALGHNTPFLHMGLAHNYNDSHMAINALRDIGYEIALDRMPRSLGPLIFVFTGSGNVS 234

Qy    190 QGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRD--IVSHKDP 247
          |||:|:|: ||| :||   ||:: |   |:|          :|||||| :|  :  |  |
Db    235 QGARELFEHLPHEYVDVATLPKVAQKGQLNK----------VYGCVVTRKDHFVPKHGGP 284

Qy    248 TRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLM------ 301
             ||| ::   |: ||  |   ||||||||:| :||: : | |: :   : |:
Db    285 ---FDKKEFEQFPDRYTSKFATEIAPYASVIINGVYWDAQSPRLITIPDAKNLLTPVQRY 341

Qy    302 -ETGCP-----LVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGA-GVV 354
             |||     |: :|||: | |||:||: : |:|::||  ||   |:  |  : | | :
Db    342 DTPGCPTLPHHLIALCDISADPGGSVEFMRECTTIDKPFAIYDADFNTSSDSFDAASGCL 401

Qy    355 CLAVDILPTEFSKEASQHFGNILSRLVASLASVKQP-----AELPSYLRRACIAHAGRLT 409
            ::| :| :   ||:: |||:|   :  | :          :  : :: | |   |:||
Db    402 VCSIDNMPAQMPIEATEQFGNLLYPWLWDLLNTSNDQHFDRLQCRTEIKNAIITDQGKLT 461

Qy    410 PLYEYIPRMRNTMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFH 469
          | :||| ::|      | | :  :
Db    462 PNFEYIAQLRKDKAASASANSRVM------------------------------------ 485

Qy    470 LVRCEVGQSTDDMSYSELEVGADDTATLDKIIDSLTSLANEHGGDHDAGQEIELALKIGK 529
                                                    ||
Db    486 ------------------------------------------GG---------------- 487

Qy    530 VNEYETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLY 589
                  | ||   ::|:|||| |  | |:| :   |             ::: ||:
Db    488 --------TTDK---RVLLLGAGMVSGPFADFYSKQKD-------------VNLTVATES 523

Qy    590 QKDAEETVDGIENTTATQLDVA-DIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKH 648
          |:| :       |  :  :|:| :  ::  |: : ::|:|||| :||  :| :||  ::
Db    524 QRDGQRLCTS-PNIQSVVVDIARESHTMERLIREHDLVVSLLPFNFHPLVAKMCISNQRD 582

Qy    649 MVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSY 708
          |||:|||   :  | :||||| |||: | ||||||||:::|:  |:     |:| :| |:
Db    583 MVTSSYVSPELEALDKAAKDADVTIMNEAGLDPGIDHMLAMECFDDIKEHGGRITSFESF 642

Qy    709 CGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRE 768
          |||||:|  ::||| |||||:| | | :  ||| |   |: : |   :: ::   : :
Db    643 CGGLPAPEWSDNPLRYKFSWSPKGVLTALMNPAKYLKNGKIVEVPSGSVVDNL--IDIDF 700

Qy    769 LPAFALEHLPNRNSLIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAANHPL 828
          :|   |   |||:|  | |:||:  :  || | | ||:|| : :  |   |     |
Db    701 MPGLNLIGFPNRDSTKYSDVYGLGNDCKTIIRGTLRYQGFVDTVKALHSVGLLSGDNIDS 760

Qy    829 LQDTSRP--TYKGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKT 886
                |  |:|    ||: ::|  |  |||     : | |   :        :::  |
Db    761 FTSNIGPDLTWK----ELIASLS--NQKLDI-----FPDSLRHII-------EEKVGGKG 802

Qy    887 VKTIKFLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPA 946
          :  :: |||  :  : :   :| | : | : : :|:  :| |:|:|:|::  : | |  :
Db    803 LSALENLGLFSDKVVDRH-GTPIDTLAQYLAKILAFKDHESDLVVLNHDIGAQLPGGN-S 860

Qy    947 EKHQATLLEFGKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPAL 1006
          |:|: :|:::|   |    :||| |||   || : ::| |::|  |: ||:  |:| |||
Db    861 ERHRISLVQYG---NPNGFSAMARTVGYTTAIVSHMVLNNEIQRAGIQRPILKEVYRPAL 917

Qy   1007 EILESSGI 1014
          : |   ||
Db    918 KRLRDFGI 925


RESULT   4
T40337
probable saccharopine dehydrogenase (NADP+, L-glutamate-forming) (EC 1.5.1.10) (simil
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Feb-2000
C;Accession: T40337
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.; Pohl, T.
submitted to the EMBL Data Library, March 1998
A;Reference number: Z21921
A;Accession: T40337
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-450 <LYN>
A;Cross-references: EMBL:AL022244; PIDN:CAA18292.1; GSPDB:GN00067; SPDB:SPBC3B8.03
A;Experimental source: strain 972h-; cosmid c3B8
C;Genetics:
A;Gene: SPDB:SPBC3B8.03
A;Map position: 2
C;Keywords: oxidoreductase


  Query Match             13.5%;  Score 716;  DB 2;  Length 450;
  Best Local Similarity   35.1%;  Pred. No. 2e-40;
  Matches  168;  Conservative   93;  Mismatches  182;  Indels   36;  Gaps   11;

Qy    544 PKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGIENT 603
          | ||:||:| |  |  |:|:                : :: ||      ||  ::|| |:
Db      2 PSILLLGSGFVAHPTLEYLSR-------------RKENNITVACRTLSKAEAFINGIPNS 48

Qy    604 TATQLDVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLS 663
           |  ||| |  :|   ||: :: |||:| ::|| :    |:  ||: | |||:  |: |
Db     49 KAIALDVNDEAALEKAVSEHDLTISLIPYTYHATVMKAAIKHGKHVCTTSYVNPKMAELE 108

Qy    664 QAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLA 723
          :||  ||   : |:|:||||||| ::| |:| |   ||||:| |||||||:|  :||||
Db    109 EAAIKAGSICMNEIGVDPGIDHLYAIKTIEEVHKAGGKIKSFLSYCGGLPAPEDSNNPLG 168

Qy    724 YKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSL 783
          |||||:  | | : :| | :   |: : :|| :| |:||   :   | :|    |||:|
Db    169 YKFSWSSRGVLLALRNSAKFYENGKLVEIDGKDLMETAKPYFI--YPGYAFVCYPNRDST 226

Qy    784 IYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFD-AANHPLLQDTSRPTYKGFLD 842
          :| : | |  || || | | ||:|| | :  |   || |  |   |  :     :|
Db    227 VYQERYQI-PEAETIIRGTLRYQGFPEFIHCLVDMGFLDETAQEYLSPEAPALPWK---- 281

Qy    843 ELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHEETQI- 901
          |:   :        |:|    : ||| ::  :   |:::   : :  :|:||:     :
Db    282 EVTARV--------IKAESSSEADLIKKISSIHKFKDEDDKKRILNGLKWLGMFSSKPVT 333
```

```
Qy    902 PKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEFGKVEN 961
          |:|  :| | :|  :|: | |   |:||::| |: |||  :|:  :    |||::| | |
Db    334 PRG--NPLDTLCATLEELMQYEEGERDMLILQHKFEVETKEGK-RQTRTCTLLDYG-VPN 389

Qy    962 GRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSGIKLVEKV 1020
          |   |:||  ||:|  :    :|   : | ||: |   ::  | :: |   ||:| |::
Db    390 G--YTSMAKLVGVPCGVATQQILDGVINTPGVLAPNDMKLCGPLIDTLAKEGIRLEEEI 446

RESULT   5
S41937
saccharopine dehydrogenase (NADP+, L-glutamate-forming) (EC 1.5.1.10) - yeast (Saccharom
N;Alternate names: protein N3461; protein YNR050c
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 27-Jan-1995 #text_change 20-Jun-2000
C;Accession: S41937; S63381
R;Feller, A.
submitted to the EMBL Data Library, January 1994
A;Reference number: S41936
A;Accession: S41937
A;Molecule type: DNA
A;Residues: 1-446 <FEL>
A;Cross-references: EMBL:X77363; NID:g453185; PID:g453186
R;Pohl, T.M.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S63346
A;Accession: S63381
A;Molecule type: DNA
A;Residues: 1-446 <POH>
A;Cross-references: EMBL:Z71665; NID:g1302563; PID:g1302564; MIPS:YNR050c
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:LYS9; LYS13
A;Cross-references: SGD:S0005333; MIPS:YNR050c
A;Map position: 14R
C;Keywords: NADP; oxidoreductase


  Query Match              12.7%;  Score 672.5;  DB 2;  Length 446;
  Best Local Similarity    35.1%;  Pred. No. 1.7e-37;
  Matches  170;  Conservative   82;  Mismatches  183;  Indels   49;  Gaps    13;

Qy    543 GPKILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGIEN 602
          |  :|:||:| | :|  : ||:              | |:| ||     :|:
Db      2 GKNVLLLGSGFVAQPVIDTLAA--------------NDDINVTVACRTLANAQALAKP-SG 47

Qy    603 TTATQLDVADIGSLSDLVSQEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNL 662
          : |  ||| |  :|  :::  :|||||:| :||  :    |  |  :||:||:  ::  |
Db     48 SKAISLDVTDDSALDKVLADNDVVISLIPYTFHPNVVKSAIRTKTDVVTSSYISPALREL 107

Qy    663 SQAAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPL 722
                ||:|:: |:||||||||| ::| ||| |   ||:|:| |||||||:|  ::|||
Db    108 EPEIVKAGITVMNEIGLDPGIDHLYAVKTIDEVHRAGGKLKSFLSYCGGLPAPEDSDNPL 167

Qy    723 AYKFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNS 782
           |||||:  | | : :| | |   |:   |   :|  :||   :   | :|    |||:|
Db    168 GYKFSWSSRGVLLALRNSAKYWKDGKIETVSSEDLMATAKPYFI--YPGYAFVCYPNRDS 225

Qy    783 LIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFF-DAANH----PLLQDTSRPTY 837
           :: ||| |  || |: | | ||:|| | :  |   |   | ||     |:  : :   |
Db    226 TLFKDLYHI-PEAETVIRGTLRYQGFPEFVKALVDMGMLKDDANEIFSKPIAWNEALKQY 284

Qy    838 KGFLDELLNNISTINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHE 897
           |                   |     :|||| :      |: |   : :    :|||
Db    285 LG--------------------AKSTSKEDLIASIDSKATWKDDEDRERILSGFAWLGLFS 325

Qy    898 ETQI-PKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEF 956
          : :| |:|  :  | :| |:|: | |  ||:|||:| |:  :|: ||   |   :||:::
Db    326 DAKITPRG--NALDTLCARLEELMQYEDNERDMVVLQHKFGIEWADG-TTETRTSTLVDY 382

Qy    957 GKVENGRSTTAMALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPAL-EILESSGIK 1015
          |||  | |  :|| ||| | ||    :|   ::  |:: |  |||  | : |: :  ||
Db    383 GKV-GGYS--SMAATVGYPVAIATKFVLDGTIKGPGLLAPYSPEINDPIMKELKDKYGIY 439
```

```
Qy   1016 LVEK 1019
          | ||
Db    440 LKEK 443

RESULT   6
T07843
probable lysine-ketoglutarate reductase / saccharopine dehydrogenase (EC 1.5.1.-) - r
N;Contains: lysine-ketoglutarate reductase; saccharopine dehydrogenase
C;Species: Brassica napus (rape)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
C;Accession: T07843
R;Deleu, C.; Coustaut, M.; Niogret, M.F.; Larher, F.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z16169
A;Accession: T07843
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-177 <DEL>
A;Cross-references: EMBL:AF042184; NID:g2809206; PIDN:AAB97685.1; PID:g2809207
A;Experimental source: cv oleifera
C;Keywords: oxidoreductase


  Query Match              11.5%;  Score 608;  DB 2;  Length 177;
  Best Local Similarity    68.6%;  Pred. No. 8.5e-34;
  Matches  118;  Conservative   24;  Mismatches   26;  Indels    4;  Gaps     2;

Qy    570 TY-GVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQEVVIS 628
          || | :  |   : ||||||| |||:||:|:    | ||||:|  ||   ||:|:||:|
Db      9 TYXGGEQRD---VRVIVASLYLKDAKETVEGMPEVEAVQLDVSDSESLLKYVSEVDVVLS 65

Qy    629 LLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMS 688
          ||||| ||::|  |||||||::||||||: | | : || ||:||| ||||||||||:|:
Db     66 LLPASCHASVAKTCIELKKHLITASYVDDETSGLHEKAKHAGITILGEMGLDPGIDHMMA 125

Qy    689 MKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNP 740
          ||||:||| ||||:|:|||||||||||||||||||||||||||||:|:|:||
Db    126 MKMINEAHIRKGKVKSFTSYCGGLPSPAAANNPLAYKFSWNPAGAIRAGRNP 177


RESULT   7
A71176
hypothetical protein PH1688 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: A71176
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophili
A;Reference number: A71000; MUID:98344137
A;Accession: A71176
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-352 <KAW>
A;Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA30800.1; PID:g3258117
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBa
C;Genetics:
A;Gene: PH1688


  Query Match               4.1%;  Score 216.5;  DB 2;  Length 352;
  Best Local Similarity    23.2%;  Pred. No. 6.9e-07;
  Matches   89;  Conservative   63;  Mismatches  149;  Indels   83;  Gaps    14;

Qy    545 KILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGIENTT 604
          |:|||||| : |  |  |                |:  | :  :  :: |:     |  |
Db      5 KVLILGAGNIGRAIAWDL---------------KDEFDVYIGDVNNENLEKVK---EFAT 46

Qy    605 ATQLDVADIGSLSDLVSQEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQ 664
```

```
                ::| ::   | ::: : |:||  ||           |: |  ||  |:: |:   |
Db         47 PLKVDASNFDKLVEVMKEFELVIGALPGFLGFKSIKAAIKSKVDMVDVSFMPENPLELRD 106

Qy        665 AAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAY 724
               |: | |||: : |  ||: :::  ::  |   ::| |     | ||||       || |
Db        107 EAEKAQVTIVFDAGFAPGLSNILMGRIFQELDLKEGYI-----YVGGLPKD--PKPPLYY 159

Qy        725 KFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLR--ELPAFALEHLPNRNS 782
              | :|:|   :     ||     |:   ||        |::::   |  ||  : | :
Db        160 KITWSPRDLIEEYTRPARVIRNGKVSKVDP---LSEVKKVKIGKFEFEAFISDGLRSMLE 216

Qy        783 LIYGDLYGISKEASTIYRATXRYEGFSEIMVTLSKTGFFDAAN--------HPLLQ---- 830
               |          :  :   | |: |  | :  | : |||   |         ||::
Db        217 TI---------NSERLEEWTLRWPGHLEKIKVLRELGFFKPENLDFTLRVIEPLMRYETK 267

Qy        831 --DTSRPTYKG-------FL----DELLNNISTINTDLDIEASGGYDDDLIARLLKLGCC 877
                   :   ||       ||    | : :::| :          |:   :|:|::    |
Db        268 DFSIMKVVGKGEEGEMEFFLYDEEDSMFSSMSRVT---------GFTAAIISRIVAENTC 318

Qy        878 KNKEIAVKTVKTI--KFLGLHEET 899
                      |   |  : ||: |:|
Db        319 --------TFGVIPPEILGMREDT 334

RESULT   8
A36467
saccharopine dehydrogenase (NAD+, L-lysine-forming) (EC 1.5.1.7) - yeast (Yarrowia lipol
C;Species: Yarrowia lipolytica, Candida lipolytica
C;Date: 15-Feb-1991 #sequence_revision 15-Feb-1991 #text_change 21-Jul-2000
C;Accession: A36467
R;Xuan, J.W.; Fournier, P.; Declerck, N.; Chasles, M.; Gaillardin, C.
Mol. Cell. Biol. 10, 4795-4806, 1990
A;Title: Overlapping reading frames at the LYS5 locus in the yeast Yarrowia lipolytica.
A;Reference number: A36467; MUID:90355996
A;Accession: A36467
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-369 <XUA>
A;Cross-references: GB:M34929; NID:g173260; PIDN:AAA35248.1; PID:g173262
C;Keywords: NAD; oxidoreductase


  Query Match             3.3%;  Score 172;  DB 2;  Length 369;
  Best Local Similarity   20.5%;  Pred. No. 0.00075;
  Matches   81;  Conservative   64;  Mismatches  159;  Indels   92;  Gaps   12;

Qy         18 IIVQPSTRRIHHDAQYEDAGCEISEDLS-----ECGLIIGIKQPKLQMILSDRAYAFFSH 72
              : |: |  ||  | :: | |  : |: |     |  :|||:|:   :       :  |:|
Db         37 VFVEKSPLRIFDDQEFVDVGATLVEEGSWVSAPEDRMIIGLKELPEESFPLSHEHIQFAH 96

Qy         73 THKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQRYL 132
               :| |     :| :      :|:| | :  |:|:|  |||  ||  ||
Db         97 CYKDQGGWKDVLSRFPAGNGTLYDLEFLEDDNGRRVAAFGFHAGFAG------------A 144

Qy        133 SLGYSTPFLSLGQSH----------MYPSLAAAKAAVIVVAEEIATFGLPSGICPIVFVF 182
              ::|  |   :  |:|           ||:       |  : :::|         | | |
Db        145 AIGVET--WAFQQTHPDSENLPGVSAYPN---ETELVDKIKKDLAAAVEKGSKLPTVLVI 199

Qy        183 TGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDIV 242
                :|   ||  :: : :         :||     |: :   :  :
Db        200 GALGRCGSGAIDLARKV--------GIPE----ENIIRWDMNETK--------------- 232

Qy        243 SHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLME 302
                       .||  :           : || | : :||:|  :  ||:| | | :
Db        233 ----------KGGPF-----------QEIAD-ADIFINCIYLSQPIPPFINYDLLNKETR 270

Qy        303 TGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDILP 362
                  :| |   | :    :     :|: : |    : :         |  :   ::| ||
Db        271 KLSVIVDVSADTTNPHNPVPVYTIATTFDHPTVPVETT--------AGPKLSVCSIDHLP 322

Qy        363 TEFSKEASQHFGNILSRLVASLASVKQPAELPSYLR 398
              :   :|||: |      |: ||  : |    | : |
Db        323 SLLPREASEAFS---EALLPSLLQLPQRDTAPVWTR 355
```

```
RESULT   9
T50174
probable saccharopine dehydrogenase [imported] - fission yeast (Schizosaccharomyces p
C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C;Accession: T50174
R;Zimmermann, W.; Wambutt, R.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, November 1999
A;Reference number: Z25036
A;Accession: T50174
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-368 <ZIM>
A;Cross-references: EMBL:AL133156; PIDN:CAB61467.1; GSPDB:GN00066; SPDB:SPAC227.18
A;Experimental source: strain 972h(-); cosmid c227
C;Genetics:
A;Gene: SPDB:SPAC227.18
A;Map position: 1


  Query Match             3.0%;  Score 157;  DB 2;  Length 368;
  Best Local Similarity   20.6%;  Pred. No. 0.0076;
  Matches   83;  Conservative   63;  Mismatches  156;  Indels  100;  Gaps   14;

Qy         17 RIIVQPSTRRIHHDAQYEDAGCEISEDLS-----ECGLIIGIKQ-PKLQMILSDRAYAFF 70
              :| :: |::|   | ::|  |  :  : |     :   |||:|: |:         :  |
Db         35 QITIERSSQRAFKDKEFERLGFPMVPEGSWRHAPKDAYIIGLKELPENDNSPLKHTHIQF 94

Qy         71 SHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAG----LIDFLHG 126
              :| :| |:    :| :       |:| | :  |:|:|  |||  || ||     : : |
Db         95 AHCYKNQEGWREVLSRFPAGNGLLYDLEFLQDDNGRRVAAFGYHAGFAGSAISCLVWAHQ 154

Qy        127 LGQRYLSLGYSTPFLSLGQSHMYPSLAA---AKAAVIVVAEEI-ATFGLPSGICPIVFVF 182
              |               |  :  :|::      |: |  || ::       :   | : |
Db        155 L---------------LHPNKQFPAIRPFPNEKSLVRHVARQVRLALKKNNNQYPRILVI 199

Qy        183 TGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRDIV 242
                :|    || ::          | |:   |    |      ||:
Db        200 GALGRCGTGACDL----------ASKIGIPFD-NILRWDINETKK--------------- 233

Qy        243 SHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLME 302
                  |  :  : |                     : |||:|     |    :: |
Db        234 --GGPFTEITESD---------------------IFVNCIYLSMPIPKFCTVESLN---V 267

Qy        303 TGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVV------CL 356
                  |  |||::||          :|:   |   |: :    |  :|  ||        :
Db        268 PNRKLRVVCDVSCD----------TTNPNNPIPIYNVNTTFDHPTVEVKGVTTPPPLEVI 317

Qy        357 AVDILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLR 398
              ::| |||   :|:|:| |      |: || ::|     | ::|
Db        318 SIDHLPTLLPRESSEAFS---EALIPSLLALKDVDNAPVWVR 356

RESULT  10
S48496
saccharopine dehydrogenase (NAD+, L-lysine-forming) (EC 1.5.1.7) - yeast (Saccharomyc
N;Alternate names: protein YIR034c
C;Species: Saccharomyces cerevisiae
C;Date: 31-Mar-1992 #sequence_revision 27-Jan-1995 #text_change 06-Feb-1998
C;Accession: S48496; S41936
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48496
A;Molecule type: DNA
A;Residues: 1-373 <ROW>
A;Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763379; MIPS:YIR034c
R;Feller, A.
submitted to the EMBL Data Library, January 1994
A;Reference number: S41936
A;Accession: S41936
```

```
A;Molecule type: DNA
A;Residues: 1-308,'V',310-373 <FEL>
A;Cross-references: EMBL:X77362; NID:g453183; PID:g453184
C;Genetics:
A;Gene: SGD:LYS1
A;Cross-references: SGD:S0001473; MIPS:YIR034c
A;Map position: 9R
C;Keywords: NAD; oxidoreductase; transmembrane protein
F;132-148/Domain: transmembrane #status predicted <TMM>

  Query Match             2.9%;  Score 154.5;  DB 2;  Length 373;
  Best Local Similarity   21.2%;  Pred. No. 0.012;
  Matches   85;  Conservative   60;  Mismatches  158;  Indels   97;  Gaps   14;

Qy     17 RIIVQPSTRRIHHDAQYEDAGCEIS-----EDLSECGLIIGIKQ-PKLQMILSDRAYAFF 70
          :| |: | :   :  :|  ||  |      :      :|||:|: |:         :  |
Db     35 KIYVEDSPQSTFNINEYRQAGAIIVPAGSWKTAPRDRIIIGLKEMPETDTFPLVHEHIQF 94

Qy     71 SHTHKAQKENMPLLDKILEERVSLFDYELIVGDDGKRSLAFGKFAGRAGLIDFLHGLGQR 130
          :| :| |     :| : ::   :|:| | :  | |:|  ||| :|| ||       || |
Db     95 AHCYKDQAGWQNVLMRFIKGHGTLYDLEFLENDQGRRVAAFGFYAGFAGA-----ALGVR 149

Qy    131 YLSLGYSTPFLSLGQSHM----------YPSLAAAKAAVIVVAEEIATFGLPSGICPIVF 180
                     :  |:|           ||:    || |  | ::           | |
Db    150 ---------DWAFKQTHSDDEDLPAVSPYPN---EKALVKDVTKDYKEALATGARKPTVL 197

Qy    181 VFTGVGNVSQGAQEIFKLLPHTFVDAEKLPEIFQARNLSKQSQSTKRVFQLYGCVVTSRD 240
          :   :|    || ::   :         :|:     |: |                   |
Db    198 IIGALGRCGSGAIDLLHKV--------GIPD----ANILKW-------------------D 227

Qy    241 IVSHKDPTR--QFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQ 298
          |   |: :|   ||:                   | | : :||:|  |   |  ||::|
Db    228 I---KETSRGGPFDE------------------IPQADIFINCIYLSKPIAPFTNMEKLN 266

Qy    299 QLMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAV 358
                  :| |   | :    |     :|   :|      :         |  :  :::
Db    267 NPNRRLRTVVDVSADTTNPHNPIPIYTVATVFNKPTVLVPTT--------AGPKLSVISI 318

Qy    359 DILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLR 398
          | ||:   :|||: | :    |: ||  : |    |:: |
Db    319 DHLPSLLPREASEFFSH---DLLPSLELLPQRKTAPVWVR 355

RESULT  11
F97474
hypothetical protein AGR_C_1727 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: F97474
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: F97474
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-366 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK86751.1; PID:g15155947; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_1727
A;Map position: circular chromosome

  Query Match             2.8%;  Score 148.5;  DB 2;  Length 366;
  Best Local Similarity   25.2%;  Pred. No. 0.028;
  Matches  105;  Conservative   42;  Mismatches  146;  Indels  123;  Gaps   21;

Qy    546 ILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGIENTTA 605
          |:::||| :    |  ||:  |     | |  |||       |    | | ||
Db      4 IVVIGAGNIGSAIAWMLAATGDY-RITVADRSADQ-------LANVPAHERVD------- 48

Qy    606 TQL-DVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQ 664
          |:: |:||  :|  |:     |:|  |    | ||   : :  | :  :   ||   :
Db     49 TEIVDIADRPALEALLKGKFAVLSAAPFHLTAGIAEAAVAVGTHYLDLTEDVESTRKVKA 108

Qy    665 AAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAY 724
           |: |   :: : || ||    :|:   | | ||  |: :     | ||     :| | |
Db    109 LAETAESALIPQCGLAPG---FISIVAADLA-ARFDKLDSVRMRVGALPQ--YPSNALNY 162

Qy    725 KFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSLI 784
            :|:  | :     |       | | |:|          ||  :||     |  |
Db    163 NLTWSTDGLINEYIEPC------EAI-VEG----------RLTAVPA-----LEEREEF- 199

Qy    785 YGDLYGISKEA------------------STIYRATXRYEGFSEIMVTLSK--------- 817
             | |:: ||                   |:   | || |   ||  |
Db    200 --SLDGVTYEAFNTSGGLGTLCATLEGKVRTMNYRTIRYPGHVAIMKALLNDLNLRNRRD 257

Qy    818 --TGFFDAANHPLLQD------TSRPTYKG-FLDELLNN----------------IST-- 850
               |: |    :||      |   |  | || |   |                |:|
Db    258 VLKDLFENALPGTMQDVVIVFVTVCGTRNGRFLQETYANKVYAGPVSGRMMSAIQITTAA 317

Qy    851 -INTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFL--------GLHE 897
           | | ||: | |         | : |  : :|:|:      |||        | ||
Db    318 GICTVLDLLAEGS--------LPQKGFVRQEEVALP-----KFLENRFGRYYGAHE 360

RESULT  12
AB2693
dehydrogenase Atu0946 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AB2693
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AB2693
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-366 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL41960.1; PID:g17739329; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu0946
A;Map position: circular chromosome

  Query Match             2.8%;  Score 148.5;  DB 2;  Length 366;
  Best Local Similarity   25.2%;  Pred. No. 0.028;
  Matches  105;  Conservative   42;  Mismatches  146;  Indels  123;  Gaps   21;

Qy    546 ILILGAGRVCRPAAEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGIENTTA 605
          |:::||| :    |  ||:  |     | |  |||       |    | | ||
Db      4 IVVIGAGNIGSAIAWMLAATGDY-RITVADRSADQ-------LANVPAHERVD------- 48

Qy    606 TQL-DVADIGSLSDLVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQ 664
          |:: |:||  :|  |:     |:|  |    | ||   : :  | :  :   ||   :
Db     49 TEIVDIADRPALEALLKGKFAVLSAAPFHLTAGIAEAAVAVGTHYLDLTEDVESTRKVKA 108

Qy    665 AAKDAGVTILCEMGLDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAY 724
           |: |   :: : || ||    :|:   | | ||  |: :     | ||     :| | |
Db    109 LAETAESALIPQCGLAPG---FISIVAADLA-ARFDKLDSVRMRVGALPQ--YPSNALNY 162

Qy    725 KFSWNPAGALRSGKNPAVYKFLGETIHVDGHNLYESAKRLRLRELPAFALEHLPNRNSLI 784
            :|:  | :     |       | | |:|          ||  :||     |  |
Db    163 NLTWSTDGLINEYIEPC------EAI-VEG----------RLTAVPA-----LEEREEF- 199

Qy    785 YGDLYGISKEA------------------STIYRATXRYEGFSEIMVTLSK--------- 817
             | |:: ||                   |:   | || |   ||  |
Db    200 --SLDGVTYEAFNTSGGLGTLCATLEGKVRTMNYRTIRYPGHVAIMKALLNDLNLRNRRD 257
```

```
Qy      818 --TGFFDAANHPLLQD------TSRPTYKG-FLDELLNN----------------IST-- 850
                 |: |    :||      |   |  | || |   |                |:|
Db      258 VLKDLFENALPGTMQDVVIVFVTVCGTRNGRFLQETYANKVYAGPVSGRMMSAIQITTAA 317

Qy      851 -INTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFL--------GLHE 897
             | | ||: | |         | : |  : :|:|:      |||        | ||
Db      318 GICTVLDLLAEGS--------LPQKGFVRQEEVALP-----KFLENRFGRYYGAHE 360

RESULT  13
S13507
microtubule-associated protein MAP2 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999
C;Accession: S13507
R;Marechal, D.; Delapierre, D.; Dresse, A.
Arch. Int. Physiol. Biochim. 96, 231-236, 1988
A;Title: Cloning and partial sequencing of a new rat brain specific cDNA.
A;Reference number: S13507; MUID:89334524
A;Accession: S13507
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-1825 <MAR>
A;Cross-references: EMBL:X54100; NID:g56624; PIDN:CAA38034.1; PID:g56625
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1990
C;Superfamily: microtubule-associated protein MAP2b; MAP2/tau repeat homology
C;Keywords: microtubule binding; tandem repeat
F;1667-1697/Domain: MAP2/tau repeat homology <MT1>
F;1698-1728/Domain: MAP2/tau repeat homology <MT2>
F;1729-1760/Domain: MAP2/tau repeat homology <MT3>

  Query Match              2.8%;  Score 145.5;  DB 2;  Length 1825;
  Best Local Similarity   20.9%;  Pred. No. 0.62;
  Matches  170;  Conservative  102;  Mismatches  271;  Indels  269;  Gaps   42;

Qy      240 DIVSHKDPTRQFDKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQ 299
            ||:::   |   |:|| |  |   ||   :         : |   | :       |: :|
Db      729 DILTNTSGT--MDEGDDYLPPT--TPAVEK---------IPCFPIESK----EEEDKTEQ 771

Qy      300 LMETGCPLVGVCDITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVD 359
               ||     |                 || | ||    |:|  | :     | | :| |
Db      772 AKVTGGQTTQV----------------ETSSESPF----PAKEYYKN-----GTV-MAPD 805

Qy      360 ILPTEFSKEASQHFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMR 419
             ||       ::      ||| ||:::  : |   |    | :| :
Db      806 -LPEMLDLAGTR------SRL-ASVSADAEVARRKSVPSEAVVAESS------------- 844

Qy      420 NTMIDLAPAKTNPLPDKKYSTLVSLSGHLFDKFLI---NEALDIIETAG--GSFHLVRCE 474
                 | |   :  | |  | |  :   :|:|: :   :   |    :|  |||:
Db      845 ---TGLPPVADDSQPVKPDSQLEDMGYCVFNKYTVPLPSPVQDSENLSGESGSFY----- 896

Qy      475 VGQSTD---------DMSYSELEVGA----DDTATLDKIIDSLTS-----LANEHGGDHD 516
              : ||         |:|  |::: |     |  | :|     :|     |: |   |
Db      897 --EGTDDKVRRDLATDLSLIEVKLAAAGRVKDEFTAEKEASPPSSADKSGLSREFDQDRK 954

Qy      517 AGQEIELAL-----------------KIG-KVNEYETDVTIDKGGPKILILGAGRVCRPA 558
            |  :::  |                 ::| ||  :   || ::   | ||         |
Db      955 ANDKLDTVLEKSEEHVDSKEHAKESEEVGDKVELFGLGVTYEQTSAKELITTKETAPERA 1014

Qy      559 AEFLASYPDICTYGVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSD 618
             : |:| |::         |||   : |   :|| :  :|         : |:| | :
Db     1015 EKGLSSVPEVAEVETTT-KADQGLDVAA---KKDDQSPLD---------IKVSDFGQM-- 1059

Qy      619 LVSQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMG 678
                         ||  :  ||  ||||        ||: ::  |:| ::
Db     1060 -------------ASGMSVDAGKTIELKFE------VDQQLTLSSEAPQET--------- 1091

Qy      679 LDPGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGK 738
                 |  |    |: :| : |   : |     :  |
Db     1092 -----DSFMG---IESSHVKDGAKVSETEVKEKVAKP----------------------- 1120
```

```
Qy      739 NPAVYKFLGETIH---VDGHNLYESA---KRLRLREL-PAFALEHLPNRNSLIYGDLY-- 789
                     : :|   ||    |||:   : | :  | |    :      |||  ::
Db     1121 ---------DLVHQEAVDKEESYESSGEHESLTMESLKPDEGKKETSPETSLIQDEVALK 1171

Qy      790 ---------GISKEASTI-YRATXRYEGFSEIMVTLSKTGFFDAANHPLL-QDTSRPTYK 838
                      :|:  |:|  :|  : |      : | |    :  : | :  | ::|  :
Db     1172 LSVEIPCPPPVSEADSSIDEKAEVQME-----FIQLPKEESTETPDIPAIPSDVTQPQPE 1226

Qy      839 GFLDELLNNISTINTDLDIEASGGYD------DDL-IARLLKLG--------CCKNKEIA 883
              : |          : :||| | ||      | | |  ||  |        |    :
Db     1227 AVVSE---PAEVRGEEEEIEAEGEYDKLLFRSDTLQITDLLVPGSREEFVETCPGEHKGV 1283

Qy      884 VKTVKTIK--------------FLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDM 929
            |::| ||:               || |         |  |    |  |:|  | |:|:
Db     1284 VESVVTIEDDFITVVQTTTDEGELGSH---------SVRFAAPVQPEEERRPYPHDEELE 1334

Qy      930 VLLHHEVEVEYPDGQP-----AEKHQATLLEF 956
            ||:  | : |  || |      || :    |:
Db     1335 VLMAAEAQAEPKDGSPDAPATPEKEEVPFSEY 1366

RESULT  14
A37981
microtubule-associated protein 2b - rat
N;Alternate names: MAP2b
N;Contains: microtubule-associated protein 2c (MAP2c)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 13-Aug-1999
C;Accession: A37981; B37981; S10003; S07887; S14568
R;Kindler, S.; Schulz, B.; Goedert, M.; Garner, C.C.
J. Biol. Chem. 265, 19679-19684, 1990
A;Title: Molecular structure of microtubule-associated protein 2b and 2c from rat br
A;Reference number: A37981; MUID:91060576
A;Accession: A37981
A;Molecule type: mRNA
A;Residues: 1-1830 <KIN>
A;Cross-references: GB:X51842; NID:g56620; PIDN:CAA36135.1; PID:g56621
A;Accession: B37981
A;Molecule type: mRNA
A;Residues: 1-151,1515-1830 <KI3>
R;Kindler, S.; Schwanke, B.; Schulz, B.; Garner, C.C.
Nucleic Acids Res. 18, 2822, 1990
A;Title: Complete cDNA sequence encoding rat high and low molecular weight MAP2.
A;Reference number: S10003; MUID:90251471
A;Accession: S10003
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-1830 <KI2>
A;Cross-references: EMBL:X51842; NID:g56620; PIDN:CAA36135.1; PID:g56621
R;Doll, T.; Papandrikopoulou, A.; Matus, A.
Nucleic Acids Res. 18, 361, 1990
A;Title: Nucleotide and amino acid sequences of embryonic rat MAP2c.
A;Reference number: S07887; MUID:90221819
A;Accession: S07887
A;Molecule type: mRNA
A;Residues: 1-151,1515-1830 <DOL>
A;Cross-references: EMBL:X17682; NID:g56622; PIDN:CAA35667.1; PID:g56623
R;Matus, A.; Doll, T.
submitted to the EMBL Data Library, May 1990
A;Reference number: S14568
A;Accession: S14568
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-476,'H',478-486,'E',488-525,'R',527-665,'V',667-670,'K',672-872,'R',874
A;Cross-references: EMBL:X53455; NID:g57619; PIDN:CAA37535.1; PID:g57620
C;Genetics:
A;Gene: map2
C;Superfamily: microtubule-associated protein MAP2b; MAP2/tau repeat homology
C;Keywords: alternative splicing; microtubule binding; tandem repeat
F;1-1830/Product: microtubule-associated protein 2b #status predicted <M2B>
F;1-151,1515-1830/Product: microtubule-associated protein 2c #status predicted <M2C>
F;1672-1702/Domain: MAP2/tau repeat homology <MT1>
F;1703-1733/Domain: MAP2/tau repeat homology <MT2>
```

```
F;1734-1765/Domain: MAP2/tau repeat homology <MT3>

  Query Match             2.7%;  Score 141.5;  DB 2;  Length 1830;
  Best Local Similarity   20.9%;  Pred. No. 1.2;
  Matches  167;  Conservative   99;  Mismatches  267;  Indels  267;  Gaps   41;

Qy     252 DKGDYYAHPEHYTPVFHERIAPYASVIVNCMYWEKRFPPLLNMDQLQQLMETGCPLVGVC 311
           |:|| |  |   ||   :         : |   | :       |: :|   ||     |
Db     743 DEGDDYLPPT--TPAVEK---------IPCFPIESK----EEEDKTEQAKVTGGQTTQV- 786

Qy     312 DITCDIGGSIEFINKSTSIERPFFRYDPSKNSYHDDMEGAGVVCLAVDILPTEFSKEASQ 371
                           || | ||    |:|  | :     | | :| | ||       ::
Db     787 ---------------ETSSESPF----PAKEYYKN-----GTV-MAPD-LPEMLDLAGTR 820

Qy     372 HFGNILSRLVASLASVKQPAELPSYLRRACIAHAGRLTPLYEYIPRMRNTMIDLAPAKTN 431
                 ||| ||:::  : |   |    | :| :                   | |   :
Db     821 ------SRL-ASVSADAEVARRKSVPSEAVVAESS----------------TGLPPVADD 857

Qy     432 PLPDKKYSTLVSLSGHLFDKFLI---NEALDIIETAG--GSFHLVRCEVGQSTD------ 480
             | |  | |  :   :|:|: :   :   |    :|  |||:       : ||
Db     858 SQPVKPDSQLEDMGYCVFNKYTVPLPSPVQDSENLSGESGSFY-------EGTDDKVRRD 910

Qy     481 ---DMSYSELEVGA----DDTATLDKIIDSLTS-----LANEHGGDHDAGQEIELAL--- 525
              |:|  |::: |     |  | :|     :|     |: |   |  |  :::  |
Db     911 LATDLSLIEVKLAAAGRVKDEFTAEKEASPPSSADKSGLSREFDQDRKANDKLDTVLEKS 970

Qy     526 --------------KIG-KVNEYETDVTIDKGGPKILILGAGRVCRPAAEFLASYPDICT 570
                         ::| ||  :   || ::   | ||         | : |:| |::
Db     971 EEHVDSKEHAKESEEVGDKVELFGLGVTYEQTSAKELITTKETAPERAEKGLSSVPEVAE 1030

Qy     571 YGVDDHDADQIHVIVASLYQKDAEETVDGIENTTATQLDVADIGSLSDLVSQVEVVISLL 630
                  |||   : |   :|| :  :|         : |:| | :
Db    1031 VETTT-KADQGLDVAA---KKDDQSPLD---------IKVSDFGQM-------------- 1063

Qy     631 PASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLDPGIDHLMSMK 690
            ||  :  ||  ||||        ||: ::  |:| ::               |  |
Db    1064 -ASGMSVDAGKTIELKFE------VDQQLTLSSEAPQET--------------DSFMG-- 1100

Qy     691 MIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNPAVYKFLGETI 750
            |: :| : |   : |     :  |                                : :
Db    1101 -IESSHVKDGAKVSETEVKEKVAKP--------------------------------DLV 1127

Qy     751 H---VDGHNLYESA---KRLRLREL-PAFALEHLPNRNSLIYGDLY-----------GIS 792
           |   ||    |||:   : | :  | |    :      |||  ::             :|
Db    1128 HQEAVDKEESYESSGEHESLTMESLKPDEGKKETSPETSLIQDEVALKLSVEIPCPPPVS 1187

Qy     793 KEASTI-YRATXRYEGFSEIMVTLSKTGFFDAANHPLL-QDTSRPTYKGFLDELLNNIST 850
           :  |:|  :|  : |      : | |    :  : | :  | ::|  :  : |
Db    1188 EADSSIDEKAEVQME-----FIQLPKEESTETPDIPAIPSDVTQPQPEAVVSE---PAEV 1239

Qy     851 INTDLDIEASGGYD------DDL-IARLLKLG--------CCKNKEIAVKTVKTIK---- 891
              : :||| | ||      | | |  ||  |        |    :  |::| ||:
Db    1240 RGEEEEIEAEGEYDKLLFRSDTLQITDLLVPGSREEFVETCPGEHKGVVESVVTIEDDFI 1299

Qy     892 ----------FLGLHEETQIPKGCSSPFDVICQRMEQRMAYGHNEQDMVLLHHEVEVEYP 941
                      || |         |  |    |  |:|  | |:|:  ||:  | : |
Db    1300 TVVQTTTDEGELGSH---------SVRFAAPVQPEEERRPYPHDEELEVLMAAEAQAEPK 1350

Qy     942 DGQP-----AEKHQATLLEF 956
           || |      || :    |:
Db    1351 DGSPDAPATPEKEEVPFSEY 1370

RESULT  15
I39837
dnaK-type molecular chaperone - Bacillus megaterium
N;Alternate names: major heat shock protein dnaK homolog
C;Species: Bacillus megaterium
C;Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 21-Jul-2000
C;Accession: I39837
R;Sussman, M.D.; Setlow, P.L.
Nucleic Acids Res. 15, 3923, 1987
A;Title: Nucleotide sequence of a Bacillus megaterium gene homologous to the dnaK gen
A;Reference number: I39837; MUID:87231083
A;Accession: I39837
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-605 <RES>
A;Cross-references: GB:M31338; NID:g39628; PIDN:CAA68348.1; PID:g39629
C;Function:
A;Description: involved in protein folding and assembling/disassembling of protein co
C;Superfamily: heat shock protein 70
C;Keywords: ATP; molecular chaperone

  Query Match             2.6%;  Score 139.5;  DB 2;  Length 605;
  Best Local Similarity   21.9%;  Pred. No. 0.26;
  Matches  143;  Conservative   81;  Mismatches  238;  Indels  191;  Gaps   30;

Qy     428 AKTNPLPDKKYSTLVSLSGHLFDKFLINEALDIIETAGGSFHLVRCEVGQST-DDMS--- 483
           | |||      :|::|:  |:                 |: | |  |  | |  :||
Db      57 AITNP------NTIISVKRHM-----------------GTDHKVEAEGKQYTPQEMSAII 93

Qy     484 ------YSELEVGADDTATLDKIIDSLTSLAN--EHGGDHDAGQ----EIE--------- 522
                 |:|  :|   |    | : :: :  |  |     |||:    |:|
Db      94 LQHLKGYAEEYLGEPVT----KAVITVPAYFNDAERQATKDAGKIAGLEVERIINEPTAA 149

Qy     523 -LALKIGKVNEYETDVTIDKGG----PKILILGAGRVCRPAAEFLASYPDICTYGVDDHD 577
            ||  : | :| :| :  | ||      || || |       |  |:  |    | || |
Db     150 ALAYGLEKTDEDQTVLVYDLGGGTFDVSILELGDG-----VFEVRATAGD-NRLGGDDFD 203

Qy     578 ADQIHVIVASLYQKD-----------------AEETVDGIENTTATQLDVADIGSLSDLV 620
              |  :||   :::                 ||:    :   |:||: :  |   :
Db     204 QVIIDYLVAEFKKENGVDLSKDKMALQRLKDAAEKAKKDLSGVTSTQISLPFI--TAGEA 261

Qy     621 SQVEVVISLLPASFHAAIAGVCIELKKHMVTASYVDESMSNLSQAAKDAGVTILCEMGLD 680
             : : :||  | |    ||:             |: :|: : || ||||
Db     262 GPLHLEVSLSRAKFDELSAGL-------------VERTMAPVRQALKDAG---------- 298

Qy     681 PGIDHLMSMKMIDEAHARKGKIKAFTSYCGGLPSPAAANNPLAYKFSWNPAGALRSGKNP 740
                 :|   :|:               ||     |  : :  :   :|      | ||
Db     299 ------LSASELDK-----------VILVGGSTRIPAVQDAIKKETGQDP----HKGVNP 337

Qy     741 AVYKFLGETIHVDGHNLYESAKRLRLRELPAFAL---------EHLPNRNSLIYGDLYGI 791
                ||  |   |  |    | : | ::   :|           |  ||: |      :
Db     338 DEVVALGAAI--QGGVLTGDVKDVVLLDVTPLSLGIETMGGVFTKLIERNTTIPTSKSQV 395

Qy     792 SKEASTIYRATXRY--EGFSEIMVTLSKTGFFDAANHPLLQDTSRPTYKGFLDELLNNIS 849
              |:    |   :  :|   :       | |   : |       |  :|        :
Db     396 FSTAADSQTAVDIHVLQGERPMSADNKTLGRFQLTDIP-------PAPRG--------VP 440

Qy     850 TINTDLDIEASGGYDDDLIARLLKLGCCKNKEIAVKTVKTIKFLGLHEETQIPKGCSSPF 909
            |    ||: :|  :     |   ||   ||| |: |:|:   |   |  :: |
Db     441 QIEVSFDIDKNGIVN----VRAKDLG--TNKEQAI-TIKSSTGLSDDEIDRMVKEAEENA 493

Qy     910 DVICQRMEQRMAYGHNEQDMVLLHHEVEVEYPDGQPAEKHQATLLEF-GKVENGRSTTAM 968
           |   || |                 |||:     |     : || :  ||||    | |
Db     494 DADKQRKE-----------------EVELRNEADQLVFTTEKTLKDLEGKVEEAEVTKAN 536

Qy     969 ALTVGIPAAIGALLLLKNKVQTKGVIRPLQPEIYVPALEILESSGIKLVEKVE 1021
                : |||      || ::    |:  : |:     ||:::  :|| |: :
Db     537 EAKDALKAAIE-----KNDLEE---IKAKKDELQ----EIVQALTVKLYEQAQ 577


Search completed: May 31, 2002, 15:47:40
Job time: 162 sec
```